A Flexible Parametric Modelling Framework for Survival Analysis

Journal Item

A Flexible Parametric Modelling Framework for Survival Analysis

Kevin Burke

University of Limerick, Ireland

M.C. Jones

Open University, U.K.

Angela Noufaily

University of Warwick, U.K.

Summary.
We introduce a general, flexible, parametric survival modelling framework which encompasses key shapes of hazard function (constant, increasing, decreasing, up-then-down, down-then-up), various common survival distributions (log-logistic, Burr type XII, Weibull, Gompertz), and includes defective distributions (cure models). This generality is achieved using four distributional parameters: two scale-type parameters – which, respectively, relate to accelerated failure time (AFT) and proportional hazards (PH) modelling – and two shape parameters. Furthermore, we advocate "multi-parameter regression" whereby more than one distributional parameter depends on covariates – rather than the usual convention of having a single covariate-dependent (scale) parameter. This general formulation unifies the most popular survival models, allowing us to consider the practical value of possible modelling choices. In particular, we suggest introducing covariates through just one or other of the two scale parameters (covering AFT and PH models), and through a "power" shape parameter (covering more complex non-AFT/non-PH effects); the other shape parameter remains covariate-independent, and handles automatic selection of the baseline distribution. We explore inferential issues and compare with alternative models through various simulation studies, with particular focus on evidence concerning the need, or otherwise, to include both AFT and PH parameters. We illustrate the efficacy of our modelling framework using data from lung cancer, melanoma, and kidney function studies. Censoring is accommodated throughout.

Keywords: Accelerated failure time; Multi-parameter regression; Power generalised Weibull distribution; Proportional hazards.

1. Introduction

This article is concerned with both theoretical and practical aspects of parametric survival analysis with a view to providing an attractive and flexible general modelling approach to analysing survival data in areas such as medicine, population health, and disease modelling. In particular, focus will be on the choice of an appropriate flexible form for the distribution of the survival outcome and the efficient use of multi-parameter regression to understand the effects of covariates on survival.

We consider a univariate lifetime random variable, $T > 0$, the primary survival outcome, whose cumulative hazard function (c.h.f.), $H(t)$, is, atypically perhaps, modelled using a flexible parametric form which we take to be

$$H(t) = \lambda\, H_0\left((\phi t)^{\gamma}\,;\kappa\right), \qquad t > 0. \tag{1.1}$$

Here, $H_0(\cdot;\kappa)$ is an underlying c.h.f. with shape parameter κ, and $\phi > 0$, $\lambda > 0$ and $\gamma > 0$ are further parameters with the following distinct interpretations: ϕ controls the horizontal scaling of the hazard function, and is well known as the accelerated failure time parameter (also, $1/\phi$ is the usual distributional scale parameter); λ controls the vertical scaling of the hazard function, and is well known as the proportional hazards parameter; and γ is a second shape parameter which is explicitly defined as a power parameter (unlike κ which can enter in potentially more complicated ways, and might even represent a vector of parameters). Were $Y = \log(T)$ to be modelled as a location-scale distribution on $\mathbb{R}$, then $\mu = -\log\phi$ and $\sigma = 1/\gamma$ would be the location and scale of that distribution, respectively, these relationships driving our preference to specify γ as a power parameter rather than as a more general shape parameter. As will be clear in the sequel, we in fact intend that only *one* of the scale parameters be present in the model, i.e., we fix $\lambda = 1$ or $\phi = 1$ so that we consider either $H_0\left((\phi t)^\gamma;\kappa\right)$ or $\lambda H_0\left(t^\gamma;\kappa\right)$, respectively. However, we write the model in a general way (with both λ and ϕ) for the purpose of unification of sub-models.

In this article, we also propose a specific choice for $H_0(t^\gamma;\kappa)$, namely

$$H_A(t;\gamma,\kappa) = \frac{\kappa+1}{\kappa}\left\{\left(1+\frac{t^\gamma}{\kappa+1}\right)^\kappa - 1\right\}, \qquad t > 0. \tag{1.2}$$

corresponding to an adapted form of the "power generalised Weibull" (PGW) distribution introduced by Bagdonavičius and Nikulin (2002); we will use APGW to stand for "adapted PGW". This choice has some major advantages: with just two shape parameters, the full range of simplest hazard shapes, namely, constant, increasing, decreasing, up-then-down or down-then-up (and no others), are available, the parameters γ and κ controlling this through the way they control behaviour of the hazard function near zero and at infinity. Here, we use the simple descriptive terms "up-then-down" and "down-then-up" to avoid the term "bathtub-shaped", which is down-then-up but with a flat valley, the clumsy term "upside-down-bathtub-shaped", and the terms "unimodal/uniantimodal" which also encompass monotone hazards. Our adaptation of the PGW distribution also allows κ to control distributional choice within the family: for $\kappa \geq 0$, log-logistic and Burr Type XII distributions are the heaviest tailed members, Weibull distributions ($\kappa = 1$) are "central" within the family, and Gompertz-related distributions are the most lightly tailed. See Section 2 for details of this model, which also include its cure model special cases when $-1 < \kappa < 0$.

Any one or more of the four distributional parameters in model (1.1) can be made to depend, typically log-linearly, on covariates; such "multi-parameter regression" (Burke and MacKenzie, 2017) is one of the focusses of this work. Indeed, this general formulation covers the most popular survival models, e.g., the accelerated failure time (AFT) model when ϕ depends on covariates, the proportional hazards (PH) model when λ depends on covariates, and semi-parametric versions when H_0 is an unspecified function. In particular, an advantage of considering (1.1) is that one may evaluate the breadth of possible modelling choices. Our primary focus in this respect is to consider which distributional parameters should depend on covariates to assess, for example, whether an AFT model (ϕ regression) is, in general, likely to provide a superior fit when compared with a PH model (λ regression), the utility of a simultaneous AFT-PH model (simultaneous ϕ and λ regression components; Chen and Jewell (2001)) when $\kappa \neq 1$, and the merits of a shape regression component (γ or κ) in addition to the, more standard, AFT and PH components.

One might also consider whether or not non-parametric components should be introduced either for functions of covariates within the regression equations, or for the baseline c.h.f., H_0, or

both. The main reason for our focus on the core model structure rather than the development of non-/semi-parametric approaches is that, within the survival literature, there is a general over-emphasis placed on semi-parametric models – compared with other fields of statistics – to the extent that many useful parametric alternatives do not receive the attention they deserve. In particular, practitioners are often content with the "flexibility" afforded by a non-parametric baseline function without concerning themselves with the possibly inflexible structural assumptions of the model at hand. Indeed, a structurally flexible parametric framework has the potential to outperform a less flexible semi-parametric model; for example, there might be more to be gained by contemplating the extension of a PH model (λ regression) to include a γ shape regression, than by extension to a non-parametric H_0. Of course, this is not to downplay the importance of a sufficiently flexible baseline function, and our proposed choice for H_0, (1.2), is quite general as it covers a wide variety of popular survival distributions.

Lying between fully non-parametric and more traditional parametric approaches, one could also model the baseline distribution using splines with a fixed set of knots (Whittemore and Keller, 1986; Efron, 1988). Compared to standard (non/semi-parametric) survival estimators (i.e., Kaplan and Meier (1958) and Cox (1972)), splines provide an estimate of the baseline hazard function, h_0, and estimation proceeds using full likelihood so that, among other things, model selection can be carried out using the Akaike or Bayesian Information Criteria (AIC/BIC). For example, Royston (2001), Royston and Parmar (2002), and Royston and Lambert (2011) consider survival regression models of the form $g(H(t|x)) = \rho(\log t) + x^T\beta$ where $g(\cdot)$ is a link function such that $g(H) = \log H$ and $g(H) = \log\{\exp(H) - 1\}$ correspond respectively to proportional hazards and proportional odds models, $\rho(\cdot)$ is a natural cubic spline with a fixed set of knots which describes the baseline distribution, x and β are vectors of covariates and regression coefficients, respectively, and where (spline-based) time-varying effects (i.e., $x^T\beta(t)$) are also possible; see Rosenberg (1995), Kooperberg et al. (1995), and Younes and Lachin (1997) for related work. More recently, Liu et al. (2018) define so-called "penalized generalized survival models" which extend the work of Royston et al. to include roughness penalties in the estimation procedure, and a wider range of splines/smoothers; these models have been implemented in the `rstpm2` package in `R` (Clements and Liu, 2019). Besides roughness, one can also penalize non-monotonic H, albeit alternative approaches can guarantee monotonicity (McLain and Ghosh, 2013; Hothorn et al., 2018). Although spline approaches have much to commend them, the APGW model, in combination with multi-parameter regression, also creates a rich modelling framework with the comparative advantage of relatively low complexity.

The multi-parameter regression approach considered in this article aligns with the generalized additive models for location, scale, and shape (GAMLSS) framework (Rigby and Stasinopoulos, 2005; Stasinopoulos and Rigby, 2007) (which is sometimes referred to as "distributional regression" (Stasinopoulos et al., 2018b)). The development of GAMLSS has been primarily focussed on estimation algorithms for these complex models in which each distributional parameter can depend on covariates (possibly non-parametrically), may include random effects, and for a wide variety of distributions; these algorithms have been implemented in the `gamlss` package (Stasinopoulos and Rigby, 2019) in `R`. However, in general, it is unlikely that this full flexibility (e.g., *all* parameters are covariate dependent) is required in practice, and, indeed, Stasinopoulos et al. (2008, p. 2) warn that "the [GAMLSS] models ... are very flexible and therefore should be used with care". Our focus in on such careful modelling: a general model structure which unifies and extends important existing models, the consideration of which distributional parameters might be best suited to covariate modelling, and the interpretation of key sub-models. Although this article focusses on survival analysis, such considerations are, of course, important

much more generally. Interestingly, although the `gamlss.cens` add-on package (Stasinopoulos et al., 2018a) extends the estimation to censored survival data, GAMLSS does not implement key sub-models of the APGW (log-logistic, Burr, Gompertz). However, the generalized gamma is available in GAMLSS, and can produce similar shapes to the APGW (albeit we recommend the latter – see Section 2.2); it is also possible to add new distributions to the `gamlss` package (see Stasinopoulos et al. (2008, Section 4.2)).

After Section 2, in which we justify our choice of baseline distribution and develop its properties, we consider the extension to regression modelling in Section 3, including model interpretation and estimation. Then, the properties of estimation within this general framework, and further practical aspects, are explored using simulated and real data in Sections 4 and 5, respectively (including comparisons with other models in Sections 4.3 and 5.4). Finally, we close with some discussion in Section 6.

2. The Specific Model for H_0

2.1. Basic Definition and Properties

We recommend for general use the APGW distribution with c.h.f. given by (1.2) and hazard function is

$$h_A(t;\gamma,\kappa) = \gamma t^{\gamma-1}\left(1+\frac{t^\gamma}{\kappa+1}\right)^{\kappa-1}, \qquad t>0. \tag{2.3}$$

This is a tractable distribution with readily available formulae for its (unimodal) density, survivor and quantile functions also. It comes about by a particular vertical and horizontal rescaling of the original PGW distribution which has c.h.f. $H_N(t;\gamma,\kappa) = (1+t^\gamma)^\kappa - 1$ (see Bagdonaviçius and Nikulin (2002), Nikulin and Haghighi (2009) and Dimitrakopoulou et al. (2007); the $\gamma = 1$ special case of H_N is the extended exponential distribution of Nadarajah and Haghighi (2011)). This resulting APGW distribution then retains attractive shape properties of the PGW distribution's hazard function, includes important survival distributions as special and limiting cases and extends to cure models, as we now show.

First, for fixed $\gamma, \kappa > 0$,

$$h_A(t;\gamma,\kappa) \sim \gamma\, t^{\gamma-1} \text{ as } t\to 0 \quad \text{and} \quad h_A(t;\gamma,\kappa) \sim (\kappa+1)^{1-\kappa}\gamma\, t^{\kappa\gamma-1} \text{ as } t\to\infty.$$

The power parameter γ controls the behaviour of the hazard function at zero: it goes to 0 (constant) ∞ as $\gamma > (=) < 1$. As $t\to\infty$, the hazard function goes to 0 (constant) ∞ as $\kappa\gamma < (=) > 1$. In fact, the APGW hazard function joins these tails smoothly in such a way that its hazard shapes are readily shown to be as listed in Table 1. Whenever the hazard function is non-monotone, its mode/antimode is at $\{(1-\gamma)(\kappa+1)/(\kappa\gamma-1)\}^{1/\gamma}$.

Defining H_A by (1.2) allows us to identify an especially large number of special and limiting cases, many important and well known, some less so, as listed in Table 2. (For the "Weibull extension" distribution, see Chen (2000) and Xie et al. (2002).) The shapes of their hazard functions, which are also given in Table 2, reflect the general shape properties of Table 1, of course. Note that the Gompertz hazard function, $h_A(t;\gamma=1,\kappa=\infty) = \exp(t)$, including both vertical and horizontal scaling parameters is $\lambda\phi\exp(\phi t)$, and this can be reparameterized as $\lambda^*\exp(\phi t)$ to arrive at the familiar form due to Gompertz (1825); see the Supplementary Material for more Gompertz-related discussion.

It can be shown that the APGW distribution retains membership of the log-location-scale-log-concave family of distributions of Jones and Noufaily (2015) and therefore, inter alia, unimodality

Table 1. Shapes of PGW hazard functions

γ	$\kappa\gamma$	shape
1	1	constant
≤ 1	≤ 1	decreasing
≤ 1	≥ 1	down-then-up
≥ 1	≤ 1	up-then-down
≥ 1	≥ 1	increasing

Here, pairs of $\leq$'s and/or $\geq$'s include the convention "and not both equal at once".

Table 2. Special and limiting cases of APGW distributions

κ	H_A	shapes of h_A	distribution	others encompassed
0	$\log(1+t^\gamma)$	decreasing, up-then-down	log-logistic	$H_A \times \lambda \Rightarrow$ Burr type XII
1	t^γ	decreasing, constant, increasing	Weibull	$\gamma = 1 \Rightarrow$ exponential
2	$t^\gamma + \frac{1}{6}t^{2\gamma}$	decreasing, down-then-up, increasing		$\gamma = 1 \Rightarrow$ linear hazard
∞	$e^{t^\gamma} - 1$	increasing, down-then-up		$H_A \times \lambda \Rightarrow$ Weibull extension; $H_A \times \lambda, \gamma = 1 \Rightarrow$ Gompertz

of densities. We also now note, for future reference, the attractive form of the quantile function associated with H_A, namely $Q_A(u) = \{H_A(-\log(1-u); 1, 1/\kappa)\}^{1/\gamma} \equiv Q_{A1}(u; \kappa)^{1/\gamma}$.

The new adaptation can also be used to widen the family of PGW distributions by taking $-1 < \kappa < 0$. For clarity, temporarily define $\psi = \kappa + 1$ so that $0 < \psi < 1$. The APGW c.h.f. can then be written as

$$H_A(t; \gamma, \psi) = \frac{\psi}{1-\psi}\left(1 - \frac{1}{\{1 + (t^\gamma/\psi)\}^{1-\psi}}\right).$$

This corresponds to a cure model with cure probability $p_\psi \equiv \lim_{t\to\infty} \exp(-H_A(t; \gamma, \psi)) = \exp\{-\psi/(1-\psi)\}$. Since the (improper) survival function is in this case of the form $p_\psi^{1-S_C(t)}$, this cure model has an interpretation as the distribution of the minimum of a Poisson number of random variables (e.g. cancer cells, tumours), each following the lifetime distribution with survival function S_C (e.g. Tsodikov et al. (2003)); here, the Poisson parameter is $\psi/(1-\psi)$ and $S_C(t) = \{1 + (t^\gamma/\psi)\}^{\psi-1}$ is the survival function of a scaled Burr Type XII distribution. The hazard functions $h_A(t; \gamma, \psi)$ follow the shape of their $\psi \to 1$ limit — the log-logistic — being decreasing for $\gamma \leq 1$ and up-then-down otherwise.

The PGW distribution, and in slightly more complicated form the APGW distribution, exhibit interesting frailty relationships between members of the families with different values of κ. We defer consideration of these frailty links to Jones et al. (2020) where we exploit them to obtain a useful bivariate shared frailty model with PGW/APGW marginal distributions. In addition, PGW and APGW distributions are written as linear transformation models in the Supplementary Material.

2.2. *Why This Particular Choice for H_0?*

The PGW/APGW distribution shares the set of hazard behaviours listed in Table 1 with two other established two-shape-parameter lifetime distributions centred on the Weibull distribution, namely, the generalised gamma (GG) and exponentiated Weibull (EW) distributions; see Jones and Noufaily (2015). Indeed, Rubio et al. (2019) choose to perform parametric survival analysis using the EW distribution for this reason. See Figure 1 for many examples of just how similar the hazard shapes of all three distributions are; in Figure 1, we have chosen the scale parameter such that each distribution has median one, used the PGW vertical scaling and otherwise specified shape parameters $\gamma, \kappa > 0$ only so that all three hazard functions behave as $t^{\gamma-1}$ as $t \to 0$ and as $t^{\kappa\gamma-1}$ as $t \to \infty$.

Further effort to choose shape parameters to match hazard functions or other aspects of the distributions even more closely is possible and has been pursued for the EW and GG distributions by Cox and Matheson (2014) and extended to the PGW distribution (what they call the generalised Weibull distribution) by Matheson et al. (2017). Cox and Matheson (2014) state that the "agreement between the two distributions [GG and EW] in our various comparisons, both graphically and in terms of the K–L [Kullback-Leibler] distance, is striking"; after a similar K–L matching exercise, Matheson et al. (2017) state that "the survival and hazard functions of the [PGW] distribution and its matched GG are visually indistinguishable." It remains, therefore, to choose between APGW, GG and EW distributions on other grounds. The GG distribution includes the Weibull and gamma distributions as special cases and the log-normal as a limiting one; the EW distribution includes the Weibull and exponentiated exponential distributions. Thus, while each includes the Weibull distribution, only the APGW includes the log-logistic, Burr, and Gompertz distributions (somewhat more commonly-used in survival literature than, say, the gamma, log-normal, or exponentiated exponential distributions) along with a cure model,

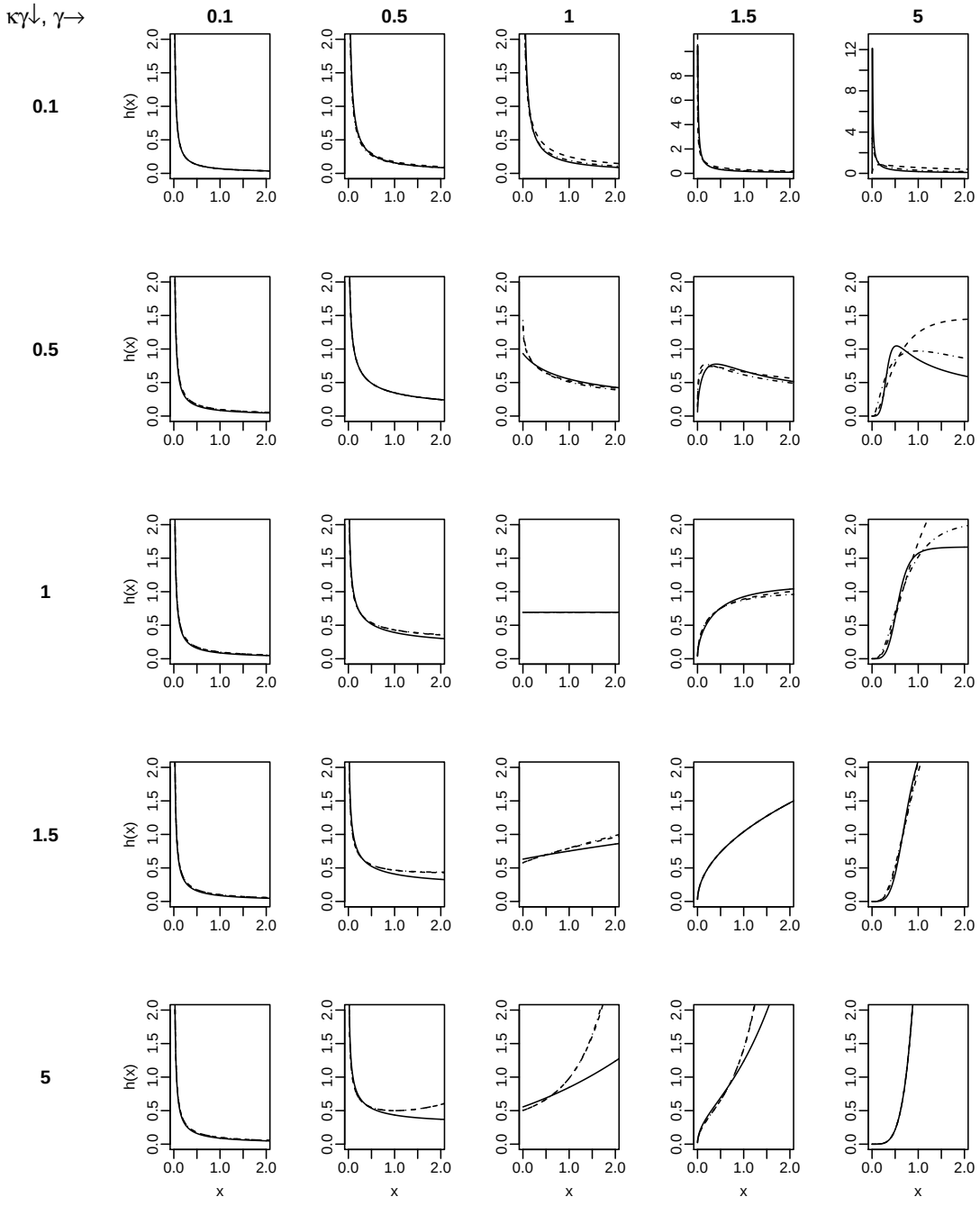

Fig. 1. Hazard functions of PGW (solid), GG (dashed), and EW (dot-dashed) distributions for the values of γ and $\kappa\gamma$ specified along the top and down the left-hand side of the figure, respectively. In each case, the scale parameter is chosen such that the median of the distributions is one. Figures on the main diagonal of the matrix of figures, in each of which the PGW, GG and EW hazard functions are identical, correspond to Weibull distributions, the figure in the centre to the exponential distribution.

where a single shape parameter, κ, selects among these distributions. We have been unable to match the number of APGW's advantageous properties as in the previous subsection by similarly adapting either the GG or EW distributions; we prefer the breadth of/difference between the wide range of distributions encompassed by the APGW distribution; and we appreciate the greater tractability of the APGW distribution both mathematically and computationally (for instance, its hazard function has a simpler form compared with the GG — which involves an incomplete gamma function — and the EW).

3. Regression

3.1. Modelling Choices

Within our proposed APGW modelling framework, there are four parameters, ϕ, λ, γ, and κ, which could potentially depend on covariates. Note that most classical modelling approaches are based on having a *single* covariate-dependent distributional parameter, which we refer to as single parameter regression (SPR), where, understandably, there is a particular emphasis on scale-type parameters, e.g., the accelerated failure time (AFT) model (ϕ regression) and the proportional hazards (PH) model (λ regression). However, in line with the flexibility of the APGW distribution itself, we also consider taking a flexible multi-parameter regression (MPR) approach in which more than one parameter may depend on covariates (cf. Burke and MacKenzie (2017), and references therein, for details of multi-parameter regression); this MPR activity might also be referred to as "distributional regression" (Stasinopoulos et al., 2018b). The most general linear APGW-MPR is, therefore, given by

$$\log(\phi) = x^T\tau, \qquad \log(\lambda) = x^T\beta, \qquad \log(\gamma) = x^T\alpha, \qquad \log(\kappa+1) = x^T\nu,$$

where log-link functions are used to respect the positivity of the parameters ϕ, λ and γ, with a slightly different link function for κ to accommodate the fact that, within our APGW, it can take values in the range $(-1, \infty)$ (see Section 2.1), $x = (1, x_1, \ldots, x_p)^T$ is a vector of covariates, and $\tau = (\tau_0, \tau_1, \ldots, \tau_p)^T$, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$, $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_p)^T$, and $\nu = (\nu_0, \nu_1, \ldots, \nu_p)^T$ are the corresponding vectors of regression coefficients. In practice, we may not necessarily have the same set of covariates appearing in all regression components, and, in our current notation, this can be handled by setting various regression coefficients to zero.

As mentioned in Section 1, we could extend the above regression specification via non-parametric regression functions of x, but this is beyond the scope of this paper and, indeed, the MPR approach is, in itself, already flexible without this added complexity. Furthermore, although the general APGW-MPR model offers the opportunity of four regression components simultaneously, this full flexibility is unlikely to be required in practice. In particular, it is well known that ϕ and λ coincide in the Weibull distribution so that only one scale parameter is needed in this case, i.e., ϕ and λ are non-identifiable when $\kappa = 1$. Moreover, our numerical studies (Sections 4 and 5) suggest that, outside of the Weibull, this is effectively much more generally true. Specifically, although ϕ and λ are theoretically identifiable in the non-Weibull cases, the parameters are nonetheless nearly non-identifiable in finite samples, which is an apparently new finding in the literature. Thus, in general, we should fix either $\phi = 1$ or $\lambda = 1$, but we would not simultaneously fix $\phi = \lambda = 1$ as a scale parameter is a core component for most statistical models. (We would expect similar considerations to apply to the work of Rubio et al. (2019), in a closely related context; they allude to numerical difficulties when $n = 1000$ [the smallest sample size they consider].)

A good practical choice is composed of the following pieces: (a) only one scale parameter (ϕ or λ) depends on covariates, while the other is fixed at one as mentioned above, (b) the γ

shape parameter may depend on covariates, and (c) the κ shape parameter is constant, i.e., only the intercept, ν_0, is non-zero in the ν vector. This choice provides a useful framework which incorporates, depending on the choice of scale regression, either an AFT (τ) or PH (β) component, allows for non-AFT/non-PH effects via the α coefficients associated with the power parameter (Section 3.2), and automatically selects the underlying baseline distribution via ν_0 from a range of popular survival distributions (Table 2) including defective distributions, i.e., cure models (Section 2.1).

3.2. Suggested models: $M(\tau, \alpha)$ and $M(\beta, \alpha)$

Let $M(\tau, \alpha)$ and $M(\beta, \alpha)$ denote the two models suggested in the previous paragraph, e.g., the latter is the model with β and α regression components, along with the shape parameter ν_0 (but where τ is a vector of zeros). More generally, beyond these two models, we will use this notation throughout the paper where the arguments of $M(\cdot)$ indicate which regression components are present in the model, the absence of either β or τ indicating that this is a vector of zeros. Irrespective of the presence of α or ν regression components in this $M(\cdot)$ notation, we assume that α_0 and ν_0 are always present since these are needed to characterise the baseline distribution and the shape of its hazard function (see Tables 1 and 2). Thus, for example, $M(\tau)$ and $M(\beta)$ are, respectively, AFT and PH models with two shape parameters (α_0 and ν_0), $M(\beta, \alpha, \nu)$ is a model which extends the suggested $M(\beta, \alpha)$ model so that the ν regression component is also present, and $M(\beta, \tau, \alpha, \nu)$ is the most general APGW-MPR model.

We first consider model $M(\tau, \alpha)$ which extends the basic AFT model, $M(\tau)$, via the incorporation of the α regression component. Now suppose that x_j is a binary covariate and let $x_{(-j)} = (1, x_1, \ldots, x_{j-1}, 0, x_{j+1}, \ldots, x_p)^T$ be the covariate vector with x_j set to zero so that we may write $x^T\tau = x_j\tau_j + x_{(-j)}^T\tau$ and $x^T\alpha = x_j\alpha_j + x_{(-j)}^T\alpha$. As this model extends the AFT model, it is natural to consider its quantile function which is given by

$$Q(u|x) = \exp(-x^T\tau)\, Q_{A1}(u;\kappa)^{\exp(-x^T\alpha)}$$

where $Q_{A1}(u) = H_A(-\log(1-u); 1, 1/\kappa)$ is the "baseline" quantile function defined in Section 2.1. We can then inspect the quantile ratio

$$QR_j(u) = \frac{Q(u|x_j = 1)}{Q(u|x_j = 0)} = \exp(-\tau_j)\, Q_{A1}(u;\kappa)^{\exp(-x_{(-j)}^T\alpha)\{\exp(-\alpha_j)-1\}}$$

where we see that α_j is the key parameter in determining the u-dependence. In particular, since $Q_{\mathrm{A1}}(u;\kappa)$ is an increasing function of u, $QR_j(u)$ increases when $\alpha_j < 0$, decreases when $\alpha_j > 0$, and is constant (i.e., the usual AFT case) when $\alpha_j = 0$. Hence, the α_j coefficient characterises the nature of the effect of the binary covariate x_j, and provides a test of the AFT property for that covariate.

Now consider the model $M(\beta, \alpha)$ which extends the PH model, $M(\beta)$, and whose hazard function is given by

$$h(t|x) = \exp(x^T\beta)\, h_A(t; \exp(x^T\alpha), \kappa)$$

where $h_A(t;\gamma,\kappa)$ is defined in (2.3). The hazard ratio for the binary covariate x_j is

$$HR_j(t) = \frac{h(t|x_j = 1)}{h(t|x_j = 0)} = \exp(\beta_j + \alpha_j)\, t^{\exp(x_{(-j)}^T\alpha)\{\exp(\alpha_j)-1\}} g(t; \alpha_j, x_{(-j)}^T\alpha, \kappa)$$

where

$$g(t;\alpha_j, x_{(-j)}^T\alpha, \kappa) = \left(\frac{t^{\exp(\alpha_j + x_{(-j)}^T\alpha)} + \kappa + 1}{t^{\exp(x_{(-j)}^T\alpha)} + \kappa + 1}\right)^{\kappa-1}.$$

Clearly, α_j characterises departures from proportional hazards as $HR_j(t)$ is a constant when $\alpha_j = 0$. For $\kappa \geq 0$, we have that $\lim_{t\to 0} HR_j(t) = 0$ and $\lim_{t\to\infty} HR_j(t) = \infty$ when $\alpha_j > 0$, while $\lim_{t\to 0} HR_j(t) = \infty$ and $\lim_{t\to\infty} HR_j(t) = 0$ when $\alpha_j < 0$. Furthermore, it can be shown that $HR_j(t)$ varies monotonically in t in the following cases: (i) $\kappa \geq 1$, or (ii) $0 < \kappa < 1$ and $\alpha_j \notin (\log\kappa, -\log\kappa)$. (We do not know about monotonicity or otherwise in the remaining cases.)

From the above we see that, within the suggested $M(\tau,\alpha)$ and $M(\beta,\alpha)$ models, the parameters play the following roles: the scale coefficients (τ or β) control the overall size of the effect where negative coefficients correspond to longer lifetimes; the α shape coefficients describe how covariate effects vary over the lifetime, i.e., permitting non-AFT and non-PH effects; and the ν_0 shape parameter characterises the baseline distribution. Note that we could, alternatively, achieve non-constant covariate effects via the ν regression component rather than the α component, i.e., using $M(\beta,\nu)$ rather than $M(\beta,\alpha)$. However, in this case, the interpretation is that such non-constant effects are due to populations which arise from structurally different distributions, rather than different shapes within a given baseline distribution. The latter is arguably more natural as it creates a separation of parameters whereas, in the former, distribution selection and non-constant covariate effects are intertwined. Of course, this is not to say that models with ν components instead of, or in combination with, α components will never be useful in practice. However, we are highlighting practical merits of the $M(\tau,\alpha)$ and $M(\beta,\alpha)$ models and, indeed, the general use of these models is motivated by the numerical studies of Sections 4 and 5.

3.3. Estimation

Consider the model formulation given in (1.1) with all four regression components, i.e., the $M(\tau,\beta,\alpha,\nu)$ model. (While we advocate the special cases $M(\tau,\alpha)$ or $M(\beta,\alpha)$, we write the estimation equations in a general form below so as to unify all potential model structures. In particular, estimation of both τ and β is not recommended in practice.) Let $\phi_i = \exp(x_i^T\tau)$, $\lambda_i = \exp(x_i^T\beta)$, $\gamma_i = \exp(x_i^T\alpha)$ and $\kappa_i = \exp(x_i^T\nu) - 1$ be the covariate-dependent distributional parameters for the ith individual with covariate vector $x_i = (1, x_{i1}, \ldots, x_{ip})^T$, and τ, β, α, and ν are the associated vectors of regression coefficients. Let $p^* = p + 1$ denote the length of x_i. Allow independent censoring by attaching to each individual an indicator δ_i which equals one if the response is observed, and zero if it is right-censored. The log-likelihood function is then given by

$$\ell(\theta) = \sum_{i=1}^{n}\left[\delta_i\left\{\log\left(\frac{\lambda_i\gamma_i z_i}{t_i}\right) + m_0(z_i;\kappa_i)\right\} - \lambda_i H_0(z_i;\kappa_i)\right]$$

where $\theta = (\tau^T, \beta^T, \alpha^T, \nu^T)^T$, $z_i = (\phi_i t_i)^{\gamma_i}$ and, in our proposed APGW case,

$$H_0(z;\kappa) = H_A(z;1,\kappa) = \frac{\kappa+1}{\kappa}\left\{\left(1 + \frac{z}{\kappa+1}\right)^{\kappa} - 1\right\},$$

$$m_0(z;\kappa) = \log h_0(z;\kappa) = (\kappa-1)\log\left(1 + \frac{z}{\kappa+1}\right).$$

As usual, the log-likelihood function can be maximised by solving the score equations

$$(U_\tau^T X,\ U_\beta^T X,\ U_\alpha^T X,\ U_\nu^T X)^T = 0_{4p^* \times 1}$$

where X is an $n \times p^*$ matrix whose ith row is x_i, $0_{4p^* \times 1}$ is a $4p^* \times 1$ vector of zeros and U_τ, U_β, U_α, and U_ν are $n \times 1$ vectors whose ith elements are as follows:

$$U_{\tau,i} = \gamma_i \left[\delta_i\{1 + z_i m_0'(z_i;\kappa_i)\} - \lambda_i z_i h_0(z_i;\kappa_i)\right]$$

$$U_{\beta,i} = \delta_i - \lambda_i H_0(z_i;\kappa_i)$$

$$U_{\alpha,i} = \delta_i[1 + \log(z_i)\{1 + z_i m_0'(z_i;\kappa_i)\}] - \lambda_i z_i \log(z_i) h_0(z_i;\kappa_i)$$

$$U_{\nu,i} = (\kappa_i + 1)\{\delta_i m_0^{(\kappa)}(z_i;\kappa_i) - \lambda_i H_0^{(\kappa)}(z_i;\kappa_i)\}$$

where

$$H_0^{(\kappa)}(z;\kappa) = \frac{\partial}{\partial\kappa} H_0(z;\kappa) = \frac{\{\kappa H_0(z;\kappa) + \kappa + 1\} m_0(z;\kappa)}{\kappa(\kappa-1)} - \frac{H_0(z;\kappa)}{\kappa(\kappa+1)} - \frac{z h_0(z;\kappa)}{\kappa+1}$$

$$m_0^{(\kappa)}(z;\kappa) = \frac{\partial}{\partial\kappa} m_0(z;\kappa) = \frac{m_0(z;\kappa)}{\kappa-1} - \frac{z m_0'(z;\kappa)}{\kappa+1}.$$

The score equations can be solved by iteratively solving the system of equations

$$\begin{pmatrix} X^T W_{\tau\tau}^{(j)} X & X^T W_{\tau\beta}^{(j)} X & X^T W_{\tau\alpha}^{(j)} X & X^T W_{\tau\nu}^{(j)} X \\ X^T W_{\tau\beta}^{(j)} X & X^T W_{\beta\beta}^{(j)} X & X^T W_{\beta\alpha}^{(j)} X & X^T W_{\beta\nu}^{(j)} X \\ X^T W_{\tau\alpha}^{(j)} X & X^T W_{\beta\alpha}^{(j)} X & X^T W_{\alpha\alpha}^{(j)} X & X^T W_{\alpha\nu}^{(j)} X \\ X^T W_{\tau\nu}^{(j)} X & X^T W_{\beta\nu}^{(j)} X & X^T W_{\alpha\nu}^{(j)} X & X^T W_{\nu\nu}^{(j)} X \end{pmatrix} \begin{pmatrix} \tau^{(j+1)} - \tau^{(j)} \\ \beta^{(j+1)} - \beta^{(j)} \\ \alpha^{(j+1)} - \alpha^{(j)} \\ \nu^{(j+1)} - \nu^{(j)} \end{pmatrix} = \begin{pmatrix} X^T U_\tau^{(j)} \\ X^T U_\beta^{(j)} \\ X^T U_\alpha^{(j)} \\ X^T U_\nu^{(j)} \end{pmatrix}$$

for $\theta^{(j+1)} = (\tau^{(j+1)^T}, \beta^{(j+1)^T}, \alpha^{(j+1)^T}, \nu^{(j+1)^T})^T$. The weight matrix $W_{\tau\tau}$ is an $n \times n$ diagonal matrix whose diagonal entries are given by $-\partial U_\tau / \partial \tau_0$, where differentiation applies elementwise to the vector U_τ; the other weight matrices are defined similarly. In the Supplementary Material, we provide the functional form for all second derivatives (i.e., the weight matrices), and `R` code implementing the above Newton-Raphson (NR) procedure – both our own implementation and another using the in-built `nlm` function. As a long-standing `R` optimiser, the latter NR implementation is likely to be more stable than our own; it can also efficiently calculate first and second derivatives numerically (but our experience suggests that providing first derivatives speeds up computations). Note that `gamlss` uses variations of the above NR estimation procedure (extended to handle non-parametric additive terms and random effects), namely: the `CG` (Cole and Green, 1992) and `RS` (Rigby and Stasinopoulos, 1996) algorithms, where the latter ignores the cross-derivatives (i.e., it only uses $W_{\tau\tau}$, $W_{\beta\beta}$, $W_{\alpha\alpha}$, and $W_{\nu\nu}$); see Rigby and Stasinopoulos (2005, Appendix B) for further details.

Note that the elements $U_{\tau,i}$, $U_{\beta,i}$ and $U_{\alpha,i}$ above are written generically so that they apply to any model of the form given in (1.1), i.e., they are not specific to the APGW case; the form of $U_{\nu,i}$, on the other hand, uses the way in which H_0 and hence $H_0^{(\kappa)}$ and $m_0^{(\kappa)}$ depend on κ. Thus, although the APGW is certainly a flexible choice (see Section 2), the first three score components

extend immediately to other baseline distributions by replacing H_0 (and, consequently, m_0 and m_0'). Estimation then proceeds once the functional form of $U_{\nu,i}$ has been re-evaluated.

Furthermore, one may, alternatively, prefer to maintain an unspecified baseline distribution, whereby ν represents an infinite-dimensional (possibly covariate-independent) parameter vector. In this case, estimation equations for the regression coefficients τ, β, and α can be based on $(U_\tau^T X,\ U_\beta^T X,\ U_\alpha^T X)$ where H_0 is replaced with an appropriate non-parametric estimator (and, similarly, for m_0 and m_0'). However, while non-parametric estimation of H_0 is reasonably straightforward (say, using a Nelson-Aalen-type estimator), it is well known that terms such as m_0 and m_0', which involve h_0 and h_0', are more difficult to estimate consistently (albeit spline and kernel smoothing could be used (Anderson and Senthilselvan, 1980; Ramlau-Hansen, 1983; Tanner and Wong, 1983; Whittemore and Keller, 1986)). We note that semi-parametric versions of the $M(\tau, \beta)$ and $M(\tau, \alpha)$ models have respectively been developed by Chen and Jewell (2001) and Burke et al. (2019). However, we are unaware of a semi-parametric $M(\tau, \beta, \alpha)$ model in the literature. In any case, such semi-parametric models are beyond the scope of the current paper and, indeed, a flexible parametric framework can cover a wide variety of applications as previously discussed in Section 1. Lastly, although we have considered estimation based on right-censored data here, because we are working parametrically, it is, of course, straightforward to deal with left-censoring, interval-censoring, and also truncation.

4. Simulation Studies

4.1. Without Covariates

Before considering estimation in the presence of covariates, we first investigate estimation in the context of the APGW model with no covariates. Thus, we simulated data from the APGW distribution parameterised in terms of the following unconstrained parameters: $\tau = \log \phi$, $\beta = \log \lambda$, $\gamma = \log \alpha$ and $\nu = \log(\kappa + 1)$. The values of the first three parameters were fixed at $\tau = 0.8$, $\beta = 0.5$, $\alpha = -0.3$, respectively, while ν was varied such that $\nu \in \{0.00, 0.22, 0.41, 0.69, 1.10, 1.61, \infty\}$ (rounded to two decimal places); note that $\nu = 0$, $\nu = 0.69$ and $\nu = \infty$ produce distributions related, respectively, to the log-logistic ($\kappa = 0$), Weibull ($\kappa = 1$) and Gompertz ($\kappa = \infty$) distributions (see Table 2). Furthermore, the sample size was fixed at 1000 and censoring times were generated from an exponential distribution such that, for each ν value, the censoring rate was fixed at approximately 30%. Within each of the seven simulation scenarios (i.e., varying ν), we fitted three different models with the aim of understanding the identifiability of parameters in a finite, but reasonably large, sample: (i) estimate all parameters, (ii) fix β at its true value, and (iii) fix β at zero. Thus, τ, α and ν are estimated in all three models. Other simulation scenarios with additional sample sizes (100 and 500) and censoring proportion (60%) can be found in the Supplementary Material, and the results are similar to what we present here (except for commensurately increased biases and standard errors).

Each scenario was replicated 1000 times, and the results are contained in Table 3. Clearly, estimation is unstable in model (i), i.e., standard errors are large. This instability arises as a consequence of attempting to estimate the scale parameters, τ and β, simultaneously. Indeed, in all cases where these parameters are estimated simultaneously, we have found that $\text{corr}(\hat{\tau}, \hat{\beta}) \approx 1$. Of course, it is well known that $\hat{\tau}$ and $\hat{\beta}$ are perfectly co-linear in the Weibull case ($\nu = 0.69$), but it is interesting to find that this extends (approximately) beyond the Weibull distribution. This appears to be a new finding in survival modelling and implies that these parameters play somewhat similar roles across a range of popular lifetime distributions (it also explains the large

standard errors observed in Table 2 of Chen and Jewell (2001)). This instability vanishes once β is fixed. In particular, when β is set to its true value of $\beta = 0.5$ (i.e., model (ii)), the estimates display very little bias. Moreover, when β is set to an incorrect value, $\beta = 0.0$ (i.e., model (iii)), $\hat{\tau}$ converges consistently to a value in the range 1.4–1.5 which compensates for the incorrect specification of β and varies smoothly with ν; the value of $\hat{\nu}$ changes somewhat from its value in model (ii), but, interestingly, $\hat{\alpha}$ does not. Furthermore, the fitted survivor curves for both models (not shown) are close to the truth, i.e., there is no reduction in quality of model fit as a consequence of fixing β to zero. Similarly (but not shown here), estimation is also stable if τ is fixed and β is estimated, and the fitted survivor curves are again close to the truth. Therefore, the choice of scale — either τ or β (fixing the other to zero) — behaves, approximately, as a model reparameterisation (which it is, exactly, in the Weibull case).

We note that, for both models (ii) and (iii), the standard error of $\hat{\nu}$ can be large when ν is large. However, this is not a concern as it is a consequence of the fact that the APGW distribution changes very little over a range of large ν values. That being said, when considering the extension to covariates, the fact that ν coefficients can grow large and have large standard errors can in turn lead to estimation instability, particularly when there are multiple covariates; therefore, keeping ν covariate-independent means that only one parameter can exhibit this behaviour, and this is straightforward to handle. By contrast, we can see that estimation of the α shape parameter is much more stable which is one reason for preferring that covariates enter through this parameter. (Another is the natural interpretation of model components as discussed in Section 3.2.)

4.2. With a Covariate

We simulated survival times according to the APGW distribution with parameters $\phi = \exp(\tau_0 + \tau_1 X)$, $\lambda = \exp(\beta)$, $\gamma = \exp(\alpha_0 + \alpha_1 X)$, and $\kappa = \exp(\nu) - 1$ where $X \sim \text{Bernoulli}(0.5)$, ν was varied according to the set $\{0.00, 0.22, 0.41, 0.69, 1.10, 1.61, \infty\}$, and the remaining parameter values were fixed at $\tau_0 = 0.8$, $\tau_1 = 0.6$, $\beta = 0.0$, $\alpha_0 = 0.2$, and $\alpha_1 = -0.5$; these values were selected to yield realistic survival times. In the notation of Section 3.2, the true model is $M(\tau, \alpha)$. As in Section 4.1, the sample size and censoring proportion were, respectively, set at 1000 and 30% (with censoring times generated from an exponential distribution). Within each of the seven scenarios (i.e., varying ν), we fitted the following three regression models: the more general $M(\tau, \beta, \alpha)$, the true $M(\tau, \alpha)$, and the misspecified $M(\beta, \alpha)$, respectively. The results, based on 1000 simulation replicates, are given in Table 4.

Mirroring the case with no covariates (Section 4.1), we find that estimation is unstable when attempting to estimate τ and β coefficients simultaneously in $M(\tau, \beta, \alpha)$, whereas estimation is stable in both the true $M(\tau, \alpha)$ and the misspecified $M(\beta, \alpha)$ models. In the latter, β coefficients converge consistently to values varying smoothly with ν. The results are broadly similar for smaller sample sizes of $n = 500$ and $n = 100$, and a higher censoring proportion of 60%, but again, of course, the biases and standard errors increase in these "lower information" scenarios (see Supplementary Material). Since, in theory (i.e., infinite samples), τ and β are only non-identifiable in the Weibull case, we also considered a much larger sample size of $n = 5000$ (see Supplementary Material); the results for $M(\tau, \beta, \alpha)$ are more stable especially for the highly-non-Weibull cases (κ near to zero or infinity), but even larger samples would be required for close-to-Weibull cases (κ close to one).

We now consider model fit by inspecting the estimated baseline survivor curves, i.e., the survivor curve for an individual with $X = 0$ which we denote by $S_0(t)$. In particular, we focus on this estimated baseline survivor function evaluated at three true quantiles, namely, $Q_0(u)$, $u =$

Table 3. Median and standard error (in brackets) of estimates

Model	ν	$\hat{\tau}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\nu}$
	0.00	1.87 (6.93)	-0.21 (4.98)	-0.26 (0.12)	0.18 (4.38)
	0.22	2.24 (11.52)	-0.54 (8.38)	-0.26 (0.15)	0.39 (7.24)
	0.41	2.83 (11.55)	-0.98 (8.58)	-0.26 (0.20)	0.49 (7.55)
(i) β : est	0.69	$*$ $*$	$*$ $*$	$*$ $*$	$*$ $*$
	1.10	0.83 (4.30)	0.38 (3.32)	-0.33 (0.14)	1.18 (6.82)
	1.61	0.53 (2.76)	0.57 (2.13)	-0.32 (0.11)	12.32 (6.47)
	∞	1.10 (1.53)	0.19 (1.21)	-0.32 (0.09)	13.04 (6.40)
	0.00	0.81 (0.15)	0.50 —	-0.30 (0.05)	0.00 (0.11)
	0.22	0.79 (0.15)	0.50 —	-0.30 (0.05)	0.23 (0.15)
	0.41	0.80 (0.15)	0.50 —	-0.30 (0.05)	0.40 (0.19)
(ii) β : true	0.69	0.79 (0.15)	0.50 —	-0.30 (0.06)	0.71 (0.31)
	1.10	0.79 (0.16)	0.50 —	-0.30 (0.06)	1.12 (1.59)
	1.61	0.80 (0.14)	0.50 —	-0.30 (0.05)	1.65 (3.72)
	∞	0.84 (0.09)	0.50 —	-0.28 (0.04)	13.05 (6.23)
	0.00	1.52 (0.12)	0.00 —	-0.29 (0.05)	0.15 (0.09)
	0.22	1.50 (0.13)	0.00 —	-0.29 (0.06)	0.33 (0.12)
	0.41	1.49 (0.13)	0.00 —	-0.30 (0.05)	0.48 (0.14)
(iii) β : zero	0.69	1.48 (0.13)	0.00 —	-0.30 (0.06)	0.68 (0.19)
	1.10	1.44 (0.13)	0.00 —	-0.31 (0.06)	0.99 (0.27)
	1.61	1.44 (0.14)	0.00 —	-0.31 (0.06)	1.27 (1.46)
	∞	1.42 (0.13)	0.00 —	-0.31 (0.06)	2.04 (4.50)

All numbers are rounded to two decimal places. For the models with fixed parameters, the "estimated" value shown is the value at which the parameter is fixed, and its standard error is then indicated by "—". Since τ and β are not simultaneously identifiable in the Weibull case ($\nu = 0.69$), all values are indicated by $*$ in Model (i) where both are to be estimated.

Table 4. Median and standard error (in brackets) of estimates

				Model(τ, β, α)				
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$	
0.00	0.98 (1.21)	0.55 (1.28)	-0.20 (1.37)	0.02 (0.70)	0.23 (0.13)	-0.50 (0.18)	0.05	(0.76)
0.22	0.98 (1.40)	0.51 (2.08)	-0.20 (1.63)	0.04 (1.61)	0.23 (0.17)	-0.50 (0.24)	0.27	(0.44)
0.41	1.00 (1.65)	0.57 (2.95)	-0.24 (1.99)	0.10 (2.48)	0.24 (0.18)	-0.51 (0.26)	0.40	(0.39)
0.69	* *	* *	* *	* *	* *	* *	*	*
1.10	0.39 (1.50)	0.08 (2.88)	0.35 (1.89)	0.21 (2.66)	0.17 (0.12)	-0.49 (0.18)	1.31	(6.88)
1.61	0.55 (1.00)	0.47 (1.62)	0.26 (1.34)	-0.04 (1.61)	0.19 (0.12)	-0.51 (0.18)	2.60	(7.04)
∞	0.88 (0.53)	0.69 (0.88)	-0.14 (0.81)	-0.06 (0.99)	0.21 (0.12)	-0.51 (0.17)	15.12	(6.46)
				Model(τ, α)				
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$	
0.00	0.80 (0.09)	0.60 (0.13)	0.00 —	0.00 —	0.21 (0.06)	-0.50 (0.06)	0.00	(0.06)
0.22	0.81 (0.09)	0.60 (0.12)	0.00 —	0.00 —	0.21 (0.06)	-0.50 (0.06)	0.22	(0.09)
0.41	0.80 (0.08)	0.59 (0.10)	0.00 —	0.00 —	0.20 (0.06)	-0.50 (0.06)	0.40	(0.11)
0.69	0.79 (0.08)	0.60 (0.10)	0.00 —	0.00 —	0.20 (0.06)	-0.50 (0.06)	0.70	(0.17)
1.10	0.80 (0.09)	0.60 (0.09)	0.00 —	0.00 —	0.20 (0.06)	-0.50 (0.06)	1.12	(0.84)
1.61	0.81 (0.08)	0.60 (0.08)	0.00 —	0.00 —	0.20 (0.07)	-0.50 (0.06)	1.59	(2.44)
∞	0.82 (0.05)	0.62 (0.06)	0.00 —	0.00 —	0.22 (0.05)	-0.50 (0.06)	13.16	(6.78)
				Model(β, α)				
ν	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\nu}_0$	
0.00	0.00 —	0.00 —	0.88 (0.11)	0.03 (0.08)	0.18 (0.05)	-0.52 (0.05)	-0.36	(0.12)
0.22	0.00 —	0.00 —	0.91 (0.11)	0.04 (0.08)	0.18 (0.06)	-0.51 (0.06)	-0.06	(0.15)
0.41	0.00 —	0.00 —	0.93 (0.13)	0.05 (0.09)	0.19 (0.06)	-0.50 (0.06)	0.21	(0.21)
0.69	0.00 —	0.00 —	0.98 (0.13)	0.06 (0.09)	0.20 (0.06)	-0.50 (0.06)	0.72	(0.35)
1.10	0.00 —	0.00 —	1.03 (0.14)	0.07 (0.11)	0.22 (0.06)	-0.50 (0.07)	1.83	(4.33)
1.61	0.00 —	0.00 —	1.18 (0.10)	0.08 (0.12)	0.27 (0.05)	-0.50 (0.07)	15.16	(6.52)
∞	0.00 —	0.00 —	1.54 (0.10)	0.10 (0.14)	0.37 (0.05)	-0.48 (0.07)	16.88	(1.28)

All numbers are rounded to two decimal places. For the models with fixed parameters, the "estimated" value shown is the value at which the parameter is fixed, and its standard error is then indicated by "—". Since τ and β are not simultaneously identifiable in the Weibull case ($\nu = 0.69$), all values are indicated by $*$ for Model(τ, β, α).

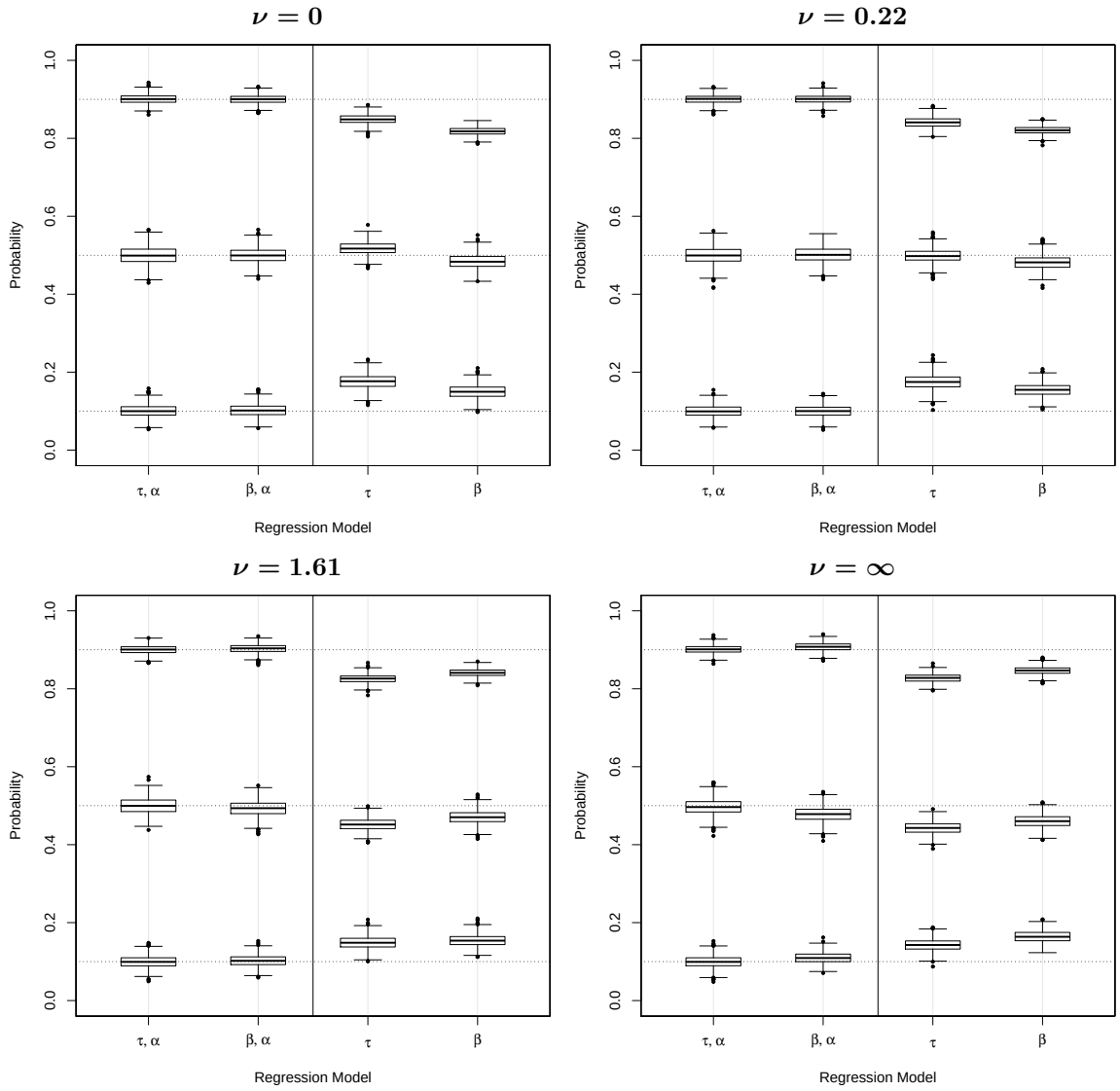

Fig. 2. Boxplots of estimated baseline survivor probabilities evaluated at the true 90th, 50th, and 10th percentile times, respectively (such that the true probabilities are 0.1, 0.5, and 0.9), vertically stacked for each of four fitted models indicated by the x-axis labels.

$0.1, 0.5, 0.9$, since $\hat{S}_0(Q_0(u))$ is an estimate of the probability $1-u$. Boxplots of these estimates over simulation replicates arising from the true model, $M(\tau, \alpha)$, and the misspecified model, $M(\beta, \alpha)$, are shown in Figure 2. We also display the estimates from two simpler (misspecified) models, $M(\tau)$ and $M(\beta)$, wherein X has been dropped from the α component (specifically, α_1 is set to zero in these simpler models). Clearly both $M(\tau, \alpha)$ and $M(\beta, \alpha)$ fit the data very well (apart from a little bias in $M(\beta, \alpha)$ when $\nu = \infty$), i.e., the choice of using a τ or β regression component does not alter the model fit much (again mirroring the findings of Section 4.1). On the other hand, when the α regression is dropped, the quality of the model fit decreases considerably as this represents a model misspecification in a much stronger sense than switching from $M(\tau, \alpha)$ to $M(\beta, \alpha)$.

4.3. Simulations with multiple covariates and model comparisons

The primary focus of the above simulation studies (Sections 4.1 and 4.2) was to show that the PH-type (β) and AFT-type (τ) parameters are close to being interchangeable in a range of distributions beyond the Weibull (where it is well-known that PH and AFT models are equivalent), and that shape regression (via α) is a fruitful activity, i.e., the $M(\tau, \alpha)$ and $M(\beta, \alpha)$ models we suggest in Section 3.2 are likely to be more useful in practice than $M(\tau, \beta)$. In this section, we focus specifically on $M(\beta, \alpha)$, investigating estimation performance with multiple covariates, and when the baseline model may be non-APGW; similar results can be obtained by focussing on $M(\tau, \alpha)$. We also compare the model fit with the generalized survival spline models of Liu et al. (2018) – available in the `rstpm2` package in `R` (Clements and Liu, 2019) – and the Cox PH model.

We simulated survival times according to a model with cumulative hazard function given by $\lambda H_0(t^\gamma)$ with parameters $\lambda = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$, $\gamma = \exp(\alpha_0 + \alpha_1 X + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4)$, and $\kappa = \exp(\nu) - 1$ where $X_1, X_2 \sim \text{Normal}(0, 0.5)$ and $X_3, X_4 \sim \text{Bernoulli}(0.5)$, and the parameter values were fixed at $\beta_0 = -1.0$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = -0.5$, $\alpha_0 = 0.2$, $\alpha_1 = \alpha_3 = 0.5$, and $\alpha_2 = \alpha_4 = 0.0$; note, therefore, that X_2 and X_4 have the PH property on account of their α coefficients being zero. The baseline cumulative hazard function was either that of an APGW model where $\nu = \log \kappa \in \{0.00, 0.69, \infty\}$, or a non-APGW model given by

$$H_0(t; \omega) = \frac{1}{2}t^2 - \frac{1}{3}t^3 + \frac{1}{4}t^4 - \frac{\omega}{\pi}\{\cos(\pi t) - 1\}$$

where $\omega \in \{0, 1, 2\}$. The hazard function for this latter model is given by $h_0(t; \omega) = t - t^2 + t^3 + \omega \sin(\pi t)$: it increases for $\omega = 0$ and has two turning points (i.e., the hazard shape is "up-then-down-then-up") for $\omega \in \{1, 2\}$. Both the sample size and censoring proportion were varied, $n \in \{500, 1000\}$ and $c \in \{30\%, 60\%\}$ (with censoring times generated from an exponential distribution), and, in all cases, we fitted the $M(\beta, \alpha)$ APGW model which is, of course, misspecified when the baseline model is given by $H_0(t; \omega)$ above.

The results, based on 1000 simulation replicates, are given in Table 5. Note that we have pooled the results for $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ into a "$\hat{\beta}_{1,2,3,4}$" column, with similarly defined "$\hat{\alpha}_{1,3}$" and "$\hat{\alpha}_{2,4}$" columns; this was done as the individual results were numerically very similar (but these individual results can be found in the Supplementary Material where we also show a smaller sample size of $n = 100$). In all cases where the truth is APGW, we find the estimates have little bias. Furthermore, the standard errors are not large relative to the average estimates (except that of $\hat{\nu}_0$ when $\nu = \infty$ which is to be expected). Thus, the estimates are recoverable for reasonable sample sizes and censoring proportions; of course, as expected, both bias and variability increase for the larger sample size and smaller censoring proportion. Interestingly,

Table 5. Median and standard error (in brackets) of estimates

			Truth: APGW					
c	n	ν	$\hat{\beta}_0$	$\hat{\beta}_{1,2,3,4}$	$\hat{\alpha}_0$	$\hat{\alpha}_{1,3}$	$\hat{\alpha}_{2,4}$	$\hat{\nu}_0$
0.3	500	0.00	-1.01 (0.14)	-0.51 (0.17)	0.22 (0.11)	0.49 (0.15)	0.01 (0.15)	0.00 (0.03)
		0.69	-1.03 (0.14)	-0.50 (0.17)	0.20 (0.10)	0.50 (0.08)	-0.01 (0.08)	0.71 (0.13)
		∞	-0.96 (0.13)	-0.51 (0.14)	0.24 (0.08)	0.50 (0.06)	0.01 (0.06)	16.62 (10.23)
	1000	0.00	-1.01 (0.10)	-0.50 (0.12)	0.21 (0.08)	0.50 (0.10)	0.00 (0.11)	0.00 (0.02)
		0.69	-1.01 (0.10)	-0.50 (0.12)	0.20 (0.07)	0.50 (0.06)	0.00 (0.05)	0.70 (0.09)
		∞	-0.98 (0.09)	-0.51 (0.10)	0.23 (0.06)	0.50 (0.04)	0.00 (0.04)	16.69 (9.88)
0.6	500	0.00	-1.02 (0.15)	-0.50 (0.19)	0.23 (0.13)	0.50 (0.16)	0.00 (0.16)	0.00 (0.06)
		0.69	-1.03 (0.17)	-0.50 (0.18)	0.21 (0.12)	0.50 (0.10)	-0.01 (0.10)	0.72 (0.20)
		∞	-0.96 (0.16)	-0.52 (0.17)	0.25 (0.10)	0.51 (0.08)	0.01 (0.08)	17.23 (21.93)
	1000	0.00	-1.00 (0.11)	-0.50 (0.13)	0.21 (0.09)	0.50 (0.11)	0.00 (0.11)	0.00 (0.04)
		0.69	-1.01 (0.11)	-0.50 (0.13)	0.20 (0.08)	0.51 (0.07)	0.00 (0.07)	0.70 (0.12)
		∞	-0.96 (0.11)	-0.51 (0.12)	0.23 (0.07)	0.51 (0.06)	0.00 (0.06)	16.45 (10.30)
			Truth: non-APGW					
c	n	ω	$\hat{\beta}_0$	$\hat{\beta}_{1,2,3,4}$	$\hat{\alpha}_0$	$\hat{\alpha}_{1,3}$	$\hat{\alpha}_{2,4}$	$\hat{\nu}_0$
0.3	500	0	-2.04 (0.21)	-0.44 (0.25)	1.07 (0.14)	0.47 (0.08)	-0.02 (0.08)	1.07 (0.24)
		1	-1.67 (0.19)	-0.47 (0.20)	0.37 (0.09)	0.49 (0.08)	-0.01 (0.08)	4.03 (10.25)
		2	-1.31 (0.16)	-0.53 (0.18)	0.17 (0.06)	0.54 (0.07)	0.03 (0.07)	18.08 (9.20)
	1000	0	-2.01 (0.15)	-0.42 (0.17)	1.07 (0.09)	0.47 (0.06)	-0.03 (0.06)	1.05 (0.16)
		1	-1.64 (0.14)	-0.46 (0.14)	0.37 (0.07)	0.49 (0.05)	-0.01 (0.06)	3.10 (8.39)
		2	-1.31 (0.11)	-0.52 (0.13)	0.16 (0.04)	0.53 (0.05)	0.02 (0.05)	17.78 (9.02)
0.6	500	0	-2.11 (0.25)	-0.45 (0.30)	1.01 (0.18)	0.47 (0.10)	-0.03 (0.10)	1.22 (1.19)
		1	-1.60 (0.22)	-0.52 (0.23)	0.35 (0.10)	0.52 (0.09)	0.00 (0.09)	3.97 (9.77)
		2	-1.19 (0.20)	-0.56 (0.20)	0.20 (0.08)	0.57 (0.08)	0.04 (0.08)	3.08 (9.26)
	1000	0	-2.04 (0.18)	-0.44 (0.20)	1.01 (0.12)	0.47 (0.07)	-0.02 (0.07)	1.16 (0.23)
		1	-1.59 (0.17)	-0.49 (0.16)	0.36 (0.07)	0.51 (0.06)	0.00 (0.06)	3.10 (8.05)
		2	-1.16 (0.15)	-0.55 (0.14)	0.20 (0.07)	0.57 (0.06)	0.04 (0.06)	2.74 (6.94)

All numbers are rounded to two decimal places. $\hat{\beta}_{1,2,3,4}$ pools results for $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ which are numerically close, and $\hat{\alpha}_{1,3}$ and $\hat{\alpha}_{2,4}$ are similarly defined.

Table 6. Survivor curve distances

	Truth:		APGW				non-APGW				
c	n	ν	$M(\beta,\alpha)$	$M(\tau,\alpha)$	GS	Cox	ω	$M(\beta,\alpha)$	$M(\tau,\alpha)$	GS	Cox
0.3	500	0.0	2.3	4.2	3.1	2.7	0	1.1	1.4	1.6	2.5
		0.7	1.2	1.4	1.4	2.9	1	2.1	2.6	2.1	4.3
		∞	1.3	1.8	2.4	4.2	2	3.5	3.8	2.5	5.7
	1000	0.0	1.6	4.0	2.3	2.1	0	0.9	1.1	1.3	2.2
		0.7	0.7	1.0	1.0	2.4	1	1.8	2.3	1.6	3.9
		∞	0.9	1.4	1.6	3.8	2	3.1	3.4	2.0	5.2
0.6	500	0.0	3.9	5.2	5.4	4.7	0	1.6	1.8	2.0	3.1
		0.7	2.5	2.8	2.6	5.3	1	2.7	3.2	3.9	5.2
		∞	2.0	2.4	3.2	4.9	2	4.2	4.6	3.6	6.7
	1000	0.0	2.8	4.7	3.7	3.7	0	1.1	1.4	1.6	2.7
		0.7	1.6	2.0	1.9	4.7	1	2.2	2.8	2.1	4.7
		∞	1.3	1.9	2.6	4.4	2	3.8	4.1	2.7	6.2

Numbers are multiplied by 100 and rounded to one decimal place. GS is the generalized survival spline model.

when the truth is non-APGW, and ignoring the intercepts (which vary from the truth to adapt to the non-APGW baseline), we see that the estimates are still not terribly biased. This suggests that covariate effects can be robust to baseline model misspecification (at least of the sort we consider here). Indeed, in the analysis of a kidney dataset (see Sections 5.3 and 5.4), we find that the hazard ratios in our model are numerically very similar to those of spline-based models even though the fit to the Kaplan-Meier survivor curves is not perfect.

In addition to $M(\beta,\alpha)$, we also fitted: $M(\tau,\alpha)$, a generalized survival spline model (GS), and a Cox PH model. For the generalized survival spline model, the cumulative hazard function was specified as $\log H(t|x) = \rho_0(\log t; df_0) + x_1\rho_1(\log t; df_1) + x_2\rho_2(\log t; df_2) + x_3\rho_3(\log t; df_3) + x_4\rho_4(\log t; df_4)$ where the $\rho(\cdot; df)$'s are natural cubic spline functions with df degrees of freedom. For simplicity, we fixed $df_0 = df_1 = df_2 = df_3 = df_4$ and selected this value by minimising BIC. A more complex estimation procedure would make use of penalized splines, but this is quite computationally intensive with multiple spline functions. (We revisit computational cost in Section 5.4.) For a given covariate profile, $x_i = (x_{1i}, x_{2i}, x_{3i}, x_{4i})^T$, the distance between the true and estimated survivor curves is

$$d_i = \int_0^{t_{\max}} \left| \hat{S}(t|x_i) - S(t|x_i) \right| dt$$

where the integral is truncated at some $t_{\max}$ since the Cox model is undefined beyond the largest observed time if it is censored, and spline models are not so well defined beyond this point either. This distance metric, averaged over individual covariate profiles and simulation replicates, and for all four fitted models, is shown in Table 6. Note that we set $t_{\max}$ equal to the largest observed time (be it censored or not) in each given simulation replicate.

Firstly, we see that overall the fit is better (distance is smaller) when the sample size is larger and the censoring proportion is smaller. When the truth is APGW, $M(\beta,\alpha)$ (the true model) provides the best fit. However, except in one case, when the truth is non-APGW with $\omega \in \{1, 2\}$, the GS (generalized survival) model does expectedly better since the "up-then-down-then-up" hazard shape is not within the APGW class of shapes; the exception to this is the case where

$c = 0.6, n = 500, \omega = 1$. The amount of improvement of GS over APGW is never very great, however, and APGW performance remains adequate. On the other hand, $M(\beta, \alpha)$ does well when $\omega = 0$ since the hazard is increasing – a shape within the APGW class. Except in the $\nu = 0.0$ cases, the $M(\beta, \alpha)$ and $M(\tau, \alpha)$ distances are fairly similar, which is not unexpected given the results of Sections 4.1 and 4.2. The Cox PH model does not do well overall since the true model is non-PH due to the α effects of X_1 and X_3.

5. Data Analysis

5.1. Lung Cancer

We now consider our modelling framework in the context of a lung cancer study which was the subject of a 1995 Queen's University Belfast PhD thesis by P. Wilkinson (see Burke and MacKenzie (2017)). This study concerns 855 individuals who were diagnosed with lung cancer between 1st October 1991 and 30th September 30 1992, who were then followed up until 30th May 1993 (approximately 20% of survival times were right-censored). The main aim of this study was to investigate differences between the following treatment groups: palliative care, surgery, chemotherapy, radiotherapy, and a combined treatment of chemotherapy and radiotherapy. In our analysis we take palliative care (which is a non-curative treatment providing pain relief) as the reference category. Note that, while various other covariates were captured for each individual, our main aim here is to explore the flexibility of our general modelling scheme in the context of the treatment variable.

As discussed in Section 3, we advocate the use of $M(\beta, \alpha)$ and $M(\tau, \alpha)$ since they offer a flexible extension of the popular PH and AFT models (i.e., $M(\beta)$ and $M(\tau)$, respectively) in which the α coefficients indicate non-PH/non-AFT effects (see Section 3.2), and where the baseline distribution is selected via the parameter $\nu_0 = \log(\kappa + 1)$. Thus, we fitted these two models, and their simpler PH and AFT counterparts, to the lung cancer data. We also fitted $M(\beta, \nu)$ and $M(\tau, \nu)$ for comparison, albeit we have argued in Section 3.2 that these are perhaps somewhat less natural. These six fitted models are summarised in Table 7.

We immediately see that the largest AIC/BIC values are associated with the simpler single component (i.e., τ and β only) models which suggests that these models are not sufficiently flexible to capture the more complex non-PH/non-AFT effects observed here. Although, in this particular application, the AFT (τ only) model has lower AIC/BIC values than that of the PH (β only) model, the fit can be greatly improved by modelling shape (either α or ν) in addition to scale. Although the best-fitting model here is $M(\beta, \nu)$, the difference is negligible compared with the models we favour, $M(\beta, \alpha)$ and $M(\tau, \alpha)$. Interestingly, these latter two models have very close AIC/BIC values, indicating that the choice of τ or β component is not at all important here (in line with the findings of Section 4.2). The use of more than two regression components did not yield further improvements in fit (models not shown), and, moreover, estimation of such models tends to be unstable — particularly, of course, those with two scale regression components (see also Section 4). Note that we have also avoided shape-only regression models, i.e., $M(\alpha)$, $M(\nu)$, and $M(\alpha, \nu)$, as, typically, models without scale components are not of interest, and, as we would expect, these models fit the current data very poorly indeed (with $\Delta_{\text{AIC}} > 600$).

We now consider the PH-APGW model, $M(\beta)$, and the two associated shape-regression extensions, $M(\beta, \alpha)$ and $M(\beta, \nu)$, in more detail. The advantage, in terms of model fit, of shape regression components is clear from Figure 3, while the $M(\beta, \alpha)$ and $M(\beta, \nu)$ models themselves are virtually indistinguishable. Table 8 displays the estimated regression coefficients. We can see that both $M(\beta)$ and $M(\beta, \alpha)$ suggest a baseline distribution which is between a log-logistic

Table 7. Summary of models fitted to lung cancer data

Model	$M(\beta)$	$M(\tau)$	$M(\beta,\alpha)$	$M(\tau,\alpha)$	$M(\beta,\nu)$	$M(\tau,\nu)$
$\dim(\theta)$	7	7	11	11	11	11
$\ell(\hat{\theta})$	-1956.5	-1943.5	-1927.0	-1926.6	-1925.9	-1930.3
Δ_{AIC}	53.1	27.2	2.2	1.5	0.0	8.9
Δ_{BIC}	34.1	8.2	2.2	1.5	0.0	8.9

$\ell(\hat{\theta})$, the value of the log-likelihood; $\dim(\theta)$, the dimension of the model, i.e., number of parameters; Δ_{AIC}, the AIC values for each model minus $\text{AIC}_{M(\beta,\nu)} = 3873.8$ (the lowest AIC in the set); Δ_{BIC}, analogous to Δ_{AIC} where the lowest BIC is $\text{BIC}_{M(\beta,\nu)} = 3926.1$.

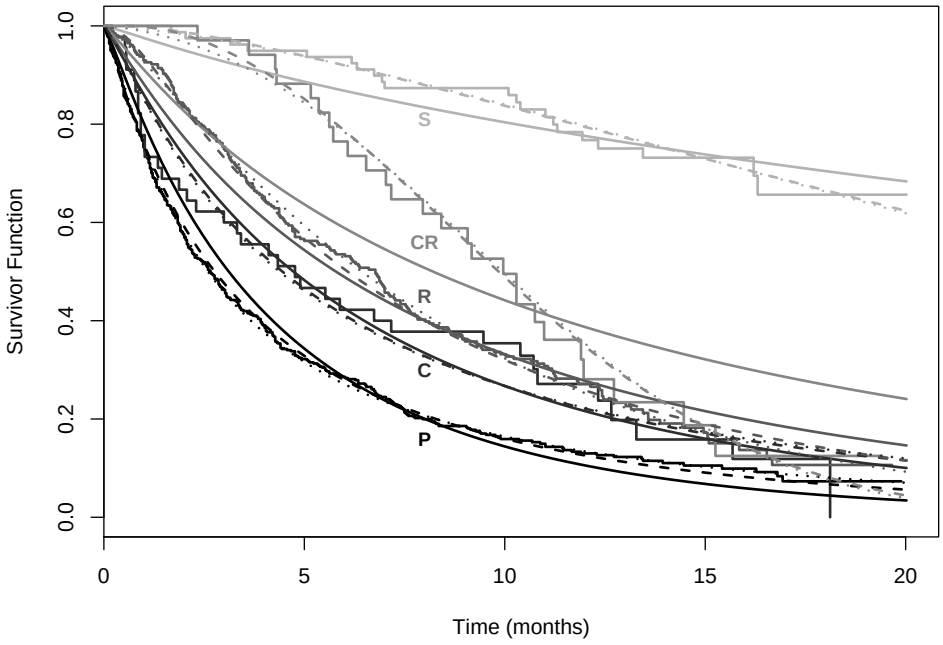

Fig. 3. Kaplan-Meier survivor curves (step, solid) for lung cancer treatment groups (P = palliative, C = chemotherapy, R = radiotherapy, CR = chemotherapy and radiotherapy, and S = surgery) with fitted curves overlayed for $M(\beta)$ (solid), $M(\beta,\alpha)$ (dashed), and $M(\beta,\nu)$ (dotted).

Table 8. Selected lung cancer models.

	Model(β)	Model(β, α)		Model(β, ν)	
	Scale	Scale	Shape	Scale	Shape
Intercept	-1.40 (0.08)	-1.13 (0.09)	0.12 (0.07)	-1.04 (0.10)	0.21 (0.06)
Palliative	0.00 —	0.00 —	0.00 —	0.00 —	0.00 —
Surgery	-2.18 (0.23)	-4.77 (0.97)	1.06 (0.28)	-3.96 (0.66)	0.55 (0.15)
Chemo	-0.38 (0.17)	-0.55 (0.33)	0.13 (0.18)	-0.60 (0.36)	0.11 (0.13)
Radio	-0.56 (0.09)	-1.46 (0.21)	0.52 (0.11)	-1.48 (0.19)	0.36 (0.06)
C+R	-0.86 (0.20)	-5.13 (0.96)	1.50 (0.22)	-3.57 (0.60)	0.82 (0.13)
$\hat{\alpha}_0$	0.15 (0.08)	$*$		0.27 (0.07)	
$\hat{\nu}_0$	0.46 (0.06)	0.35 (0.05)		$*$	

The $*$ symbol indicates that the shape parameter already appears as the intercept in the shape regression component. Standard errors in brackets.

($\nu = 0$) and a Weibull ($\nu = 0.69$), while $M(\beta, \nu)$ assumes a separate baseline distribution for each treatment group. Interestingly, in all three models, all shape parameters (ν and α) are positive which indicates that the hazards are increasing with time in each treatment group (Table 1). While all three models are in agreement when it comes to the overall effectiveness of each treatment as viewed in terms of the scale coefficients (albeit the chemotherapy effect is only statistically significant in $M(\beta)$), the positive shape coefficients in $M(\beta, \alpha)$ suggest that the effectiveness of each treatment reduces to some extent over time (see Section 3.2) – especially in the case of the combined treatment of chemotherapy and radiotherapy.

The hazard ratios for the models in Table 8 are shown in Figure 4 where those of $M(\beta, \alpha)$ and $M(\beta, \nu)$ are quite similar. They suggest that while the various treatments reduce the hazard in the first few months, their effect is weakened over time and, perhaps, even become inferior to palliative care in the longer term (however, note that very few individuals remain in the sample beyond 15 months). Clearly SPR models, such as $M(\beta)$, cannot account for covariate effects of this sort.

It is worth highlighting the fact that the basic findings here are qualitatively similar to those of Burke and MacKenzie (2017) who analysed this lung cancer dataset using $M_{\kappa=1}(\beta, \alpha)$, i.e., a Weibull MPR model. However, the framework of the current paper permits us to consider a much wider range of model structures and distributions in which $M_{\kappa=1}(\beta, \alpha)$ appears as a special case. In particular, $M(\beta, \alpha)$ from Table 8 yields a 95% confidence interval for κ, $[0.28, 0.57]$, which does not support the Weibull ($\kappa = 1$) baseline distribution. Furthermore, $\text{AIC}_{M_{\kappa=1}(\beta,\alpha)} - \text{AIC}_{M(\beta,\alpha)} = 57.6$, and we can confirm that the improvement in quality of fit is most evident in the palliative care group (which $M_{\kappa=1}(\beta, \alpha)$ does not capture so well). Thus, although the basic findings are unaltered in this particular application, the APGW MPR approach yields a better solution in which uncertainty in selecting the baseline distribution is accounted for. Of course, the APGW MPR model will readily adapt to other applications which might differ significantly (both qualitatively and quantitatively) from $M_{\kappa=1}(\beta, \alpha)$.

5.2. Melanoma

The Eastern Cooperative Oncology Group (ECOG) trial "EST 1684" was a randomized controlled trial to investigate the adjuvant (i.e., post-surgery) chemotherapy drug "IFNα-2b" in treating melanoma (Kirkwood et al., 1996). The outcome variable was relapse-free survival, i.e.,

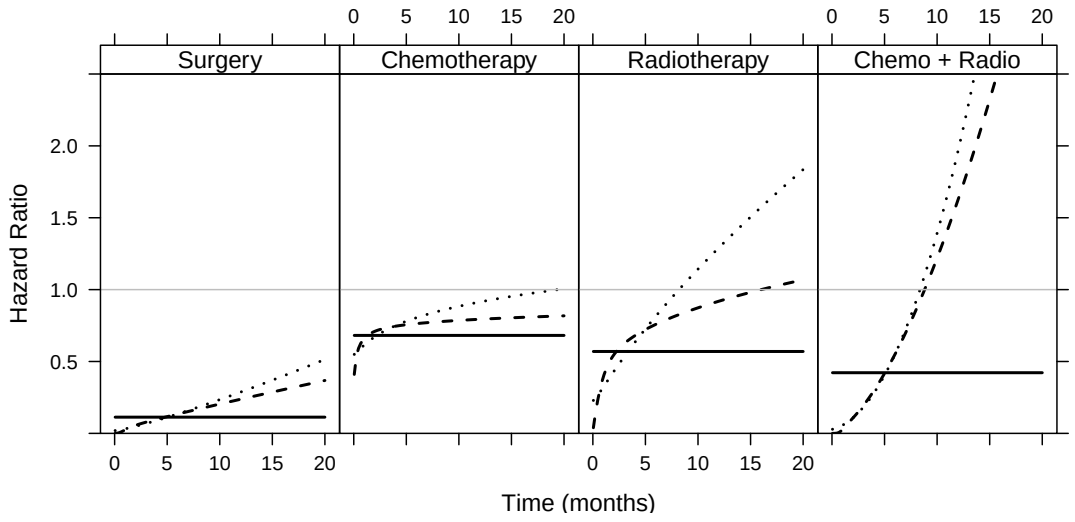

Fig. 4. Hazard ratios for each lung cancer treatment relative to palliative care for $M(\beta)$ (solid), $M(\beta, \alpha)$ (dashed), and $M(\beta, \nu)$ (dotted). Line of equality (grey) also shown.

Table 9. Summary of models fitted to melanoma data

Model	$M(\beta)$	$M(\tau)$	$M(\beta, \alpha)$	$M(\tau, \alpha)$	$M(\beta, \nu)$	$M(\tau, \nu)$
$\dim(\theta)$	4	4	5	5	5	5
$\ell(\hat{\theta})$	-368.8	-368.0	-367.7	-367.9	-368.2	-366.6
Δ_{AIC}	2.4	0.8	2.1	2.6	3.2	0.0
Δ_{BIC}	1.6	0.0	5.0	5.5	6.1	2.9

$\ell(\hat{\theta})$, the value of the log-likelihood; $\dim(\theta)$, the dimension of the model, i.e., number of parameters; Δ_{AIC}, the AIC values for each model minus $\text{AIC}_{M(\tau,\nu)} = 743.2$ (the lowest AIC in the set); Δ_{BIC}, analogous to Δ_{AIC} where the lowest BIC is $\text{BIC}_{M(\tau)} = 758.6$.

time from randomization until the earlier of cancer relapse or death. Patients were recruited to the study between 1984 and 1990, and the study ended in 1993. In total, 284 patients were recruited of which 140 were assigned to the control group, and 144 were assigned to the treatment group. This dataset is available in the `R` package `smcure` (Chao et al., 2012), and variations of it have appeared in the cure model literature (Chen et al., 1999; Ibrahim et al., 2001).

As in Section 5.1, we fitted the following models: $M(\beta)$, $M(\tau)$, $M(\beta, \alpha)$, $M(\tau, \alpha)$, $M(\beta, \nu)$, and $M(\tau, \nu)$; the results are summarised in Table 9. In this case, the two-component models do not provide a large improvement over the one-component models, and $M(\tau)$ has the lowest BIC and the second-lowest AIC ($M(\tau, \nu)$ has the lowest AIC); the AIC and BIC values for $M(\beta)$ are not much larger than for $M(\tau)$. The models $M(\beta)$, $M(\tau)$, and $M(\tau, \nu)$ are compared to the Kaplan-Meier (KM) curves in Figure 5. The fitted $M(\beta)$ and $M(\tau, \nu)$ curves are similar, and are close to the KM curves. The fitted $M(\tau)$ curves converge later in time, which, visually, look worse compared to the KM curves. However, note that there is very little data in the right tail so that converging curves are plausible when viewed with the level of uncertainty in the tail.

The parameter estimates for $M(\beta)$, $M(\tau)$, and $M(\tau, \nu)$ are shown in Table 10. Firstly

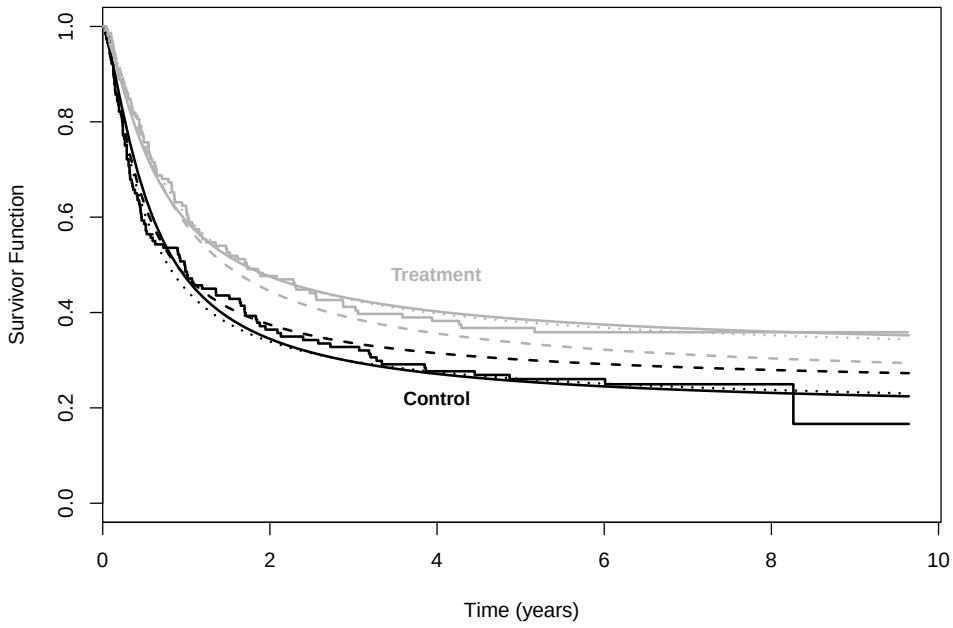

Fig. 5. Kaplan-Meier survivor curves (step, solid) for melanoma control and treatment groups with fitted curves overlayed for $M(\beta)$ (solid), $M(\tau)$ (dashed), and $M(\tau, \nu)$ (dotted).

note that the scale coefficient of treatment is negative which indicates that treatment improves survival. Furthermore, for $M(\beta)$ and $M(\tau)$, note that the parameter $\kappa = \exp(\nu_0) - 1$ is negative, and, similarly, for $M(\tau, \nu)$, the estimated κ is negative in each treatment group. Thus, all three models point towards a cure proportion. The fact that the power shape parameter, $\gamma = \exp(\alpha)$, is greater than one for all three models means that the hazard function has an up-then-down shape. This type of hazard function is commonly observed in the cure literature since the population becomes increasingly composed of cured individuals (i.e., zero hazard) over time.

When $-1 < \kappa < 0$, the APGW-MPR cure proportion is given by $\exp\{\lambda(\kappa + 1)/\kappa\} > 0$, i.e., the cure proportion depends on κ and λ. Therefore, $M(\beta)$ and $M(\tau, \nu)$ suggest that the cure proportion depends on treatment, while $M(\tau)$ suggests that it does not. The estimated cure proportions, along with 95% confidence intervals, are shown in Ta-

Table 10. Selected melanoma models.

	Model(β)		Model(τ)		Model(τ, ν)			
	Scale		Scale		Scale		Shape	
Intercept	0.60	(0.16)	0.65	(0.14)	0.67	(0.15)	-0.45	(0.05)
Control	0.00	—	0.00	—	0.00	—	0.00	—
Treatment	-0.36	(0.14)	-0.52	(0.20)	-0.53	(0.19)	-0.14	(0.08)
$\hat{\alpha}_0$	0.36	(0.07)	0.44	(0.08)			0.45	(0.08)
$\hat{\nu}_0$	-0.73	(0.11)	-0.51	(0.05)			$*$	

The $*$ symbol indicates that the shape parameter already appears as the intercept in the shape regression component. Standard errors in brackets.

Table 11. Estimated melanoma cure proportions with 95% confidence intervals

Model	Treatment	Control	Difference
$M(\beta)$	0.30 (0.21,0.39)	0.18 (0.10,0.26)	0.12 (0.03,0.21)
$M(\tau)$	0.22 (0.15,0.30)	= Treatment	—
$M(\tau,\nu)$	0.24 (0.18,0.39)	0.11 (0.09,0.27)	0.12 (-0.02,0.23)

$p_{\text{Difference}} = p_{\text{Treatment}} - p_{\text{control}}$.

ble 11. Had we fixed to a log-logistic baseline (i.e., $\kappa = 0$), the resulting $M_{\kappa=0}(\beta)$ and $M_{\kappa=0}(\tau)$ models (not shown) provide an extremely poor fit to the data. This is noteworthy as even the heaviest-tail non-cure APGW model is not supported by the data (and, of course, a Weibull baseline is worse still). The heaviness of tail here can only be supported within the APGW family by a cure model.

5.3. Kidney Function

The Graduate Entry Medical School (GEMS), University of Limerick, Ireland is currently leading a study which aims to develop Ireland's first national surveillance system for tracking kidney disease. The key measure of kidney function is the glomerular filtration rate (GFR) which is the rate at which filtered blood flows through the kidneys (mL/min/1.73m^2), and it is of interest to explore the relationship between GFR and mortality. We consider a sample of 6157 males aged 80+ who were recruited during the period 1st January 2007 to 31st December 2012, and were followed up until the earlier of death (all causes) or 31st December 2013; during the follow up period, 2692 deaths occurred. These individuals had their GFR values measured on entry to the study, and were placed into one of five categories: normal kidney function to mild loss of function (GFR $\geq$ 60), mild to moderate loss (45 $\leq$ GFR $<$ 60), moderate to severe loss (30 $\leq$ GFR $<$ 45), severe loss (15 $\leq$ GFR $<$ 30), and kidney failure (GFR $<$ 15). (These are standard categories used in the renal literature.)

Here we consider only $M(\beta)$ (a proportional hazards model) and its extension to $M(\beta,\alpha)$. Other MPR-APGW models do not fit the data significantly better as could be expected from Sections 4, 5.1, and 5.2. These fitted models are summarised in Table 12. $M(\beta)$ has a lower AIC and BIC, and, moreover, the α coefficients are not statistically significant in $M(\beta,\alpha)$; thus, the proportional hazards assumption is supported. Note that $\hat{\kappa} = \exp(\hat{\nu}_0) - 1 = 1.07$ is very close to a Weibull baseline (and, indeed, the 95% confidence interval for κ includes unity). From Figure 6 we see that $M(\beta)$ provides a reasonable fit to the data, but it is certainly not perfect. The hazard ratios (relative to the GFR $\geq$ 60 group) from this model are given, respectively, by 1.06, 1.29, 1.74, and 1.85; the increased risk of mortality with declining kidney function is clear.

Although the GFR categories above are standard in the renal literature, it is also of interest to explore the functional relationship between the original continuous GFR variable and mortality. Of course, the categorical model forms a discrete approximation to a continuous functional effect, and, from Table 12 it is clear that the GFR effect is non-linear. We investigated the GFR effect using a fractional polynomial approach (as this can be implemented straightforwardly within our optimization scheme without additional complexity). In particular we consider here a second-order fractional polynomial (higher orders did not improve AIC/BIC) which is defined as

$$F_2(x) = \begin{cases} \beta_1 x^{(a_1)} + \beta_2 x^{(a_2)}, & a_1 \neq a_2, \\ \beta_1 x^{(a_1)} + \beta_2 x^{(a_2)} \log x, & a_1 = a_2, \end{cases}$$

Table 12. Model(β) and Model(β, α) for categorical kidney function.

	Model(β)	Model(β, α)	
	Scale	Scale	Shape
Intercept	-1.42 (0.06)	-1.41 (0.06)	-0.66 (0.04)
GFR ≥ 60	0.00 —	0.00 —	0.00 —
$45 \leq$ GFR < 60	0.06 (0.05)	0.05 (0.06)	0.01 (0.05)
$30 \leq$ GFR < 45	0.25 (0.05)	0.23 (0.06)	0.04 (0.05)
$15 \leq$ GFR < 30	0.55 (0.06)	0.56 (0.07)	-0.01 (0.05)
GFR < 15	0.62 (0.12)	0.68 (0.12)	-0.15 (0.10)
$\hat{\alpha}_0$	-0.66 (0.04)	$*$	
$\hat{\nu}_0$	0.73 (0.10)	0.73 (0.10)	
$\ell(\hat{\theta})$	-6744.3	-6742.4	
AIC	13502.5	13506.7	
BIC	13549.6	13580.7	

The $*$ symbol indicates that the shape parameter already appears as the intercept in the shape regression component. Standard errors in brackets. $\ell(\hat{\theta})$ is the value of the log-likelihood.

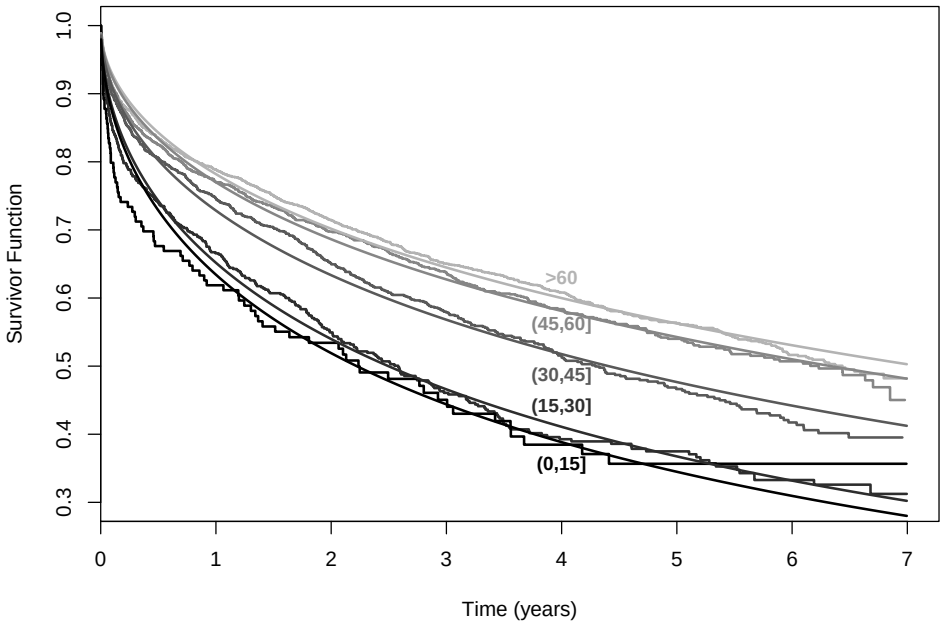

Fig. 6. Kaplan-Meier survivor curves (step, solid) for GFR groups with fitted curves overlayed for $M(\beta)$ (solid). Note: the lower limit on the y-axis is 0.3 rather than zero.

where $x^{(a)} = x^a(\log x)^{1(a=0)}$ and $1(\cdot)$ is an indicator function (Royston and Altman, 1994; Royston and Sauerbrei, 2007; Sauerbrei et al., 2007; Strasak et al., 2011). Typically, a_1 and a_2 are selected from the discrete set $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, and we follow this convention here.

We consider the GFR effect only in the β regression component since, from Table 12, there is no strong α effect. We refer to this fractional polynomial model as $M_F(\beta)$, The powers selected (based on the likelihood function) were $a_1 = a_2 = 2$, i.e., the GFR effect is modelled as $\beta_1 \text{GFR}^2 + \beta_2 \text{GFR}^2 \log \text{GFR}$. The AIC and BIC values for this model are 13489.7 and 13523.4, respectively, both lower than those of the categorical effects model, $M(\beta)$ (see Table 12). For comparison, we also fitted a model in which GFR enters linearly. This model, which we denote by $M_L(\beta)$, has much larger AIC (13514.9) and BIC (13541.8) values respectively, i.e., the linear effect is not supported (as expected from $M(\beta)$).

The hazard ratios for $M_F(\beta)$ and $M_L(\beta)$ are shown in Figure 7. The reference GFR value used is 76 as this is the mean value in the GFR ≥ 60 group; thus, the hazard ratios are broadly comparable to those of the categorical $M(\beta)$ which is also shown in Figure 7. We see that the hazard increases dramatically once the GFR value falls below 60. For comparison, we also fitted a Cox model where the GFR effect was modelled using penalized B-spines (Eilers and Marx, 1996); this can be implemented using the `pspline` function in the `survival` package in `R` (Therneau and Lumley, 2019). The spline effect is quite similar to the fractional polynomial, but is just outside the confidence interval for the largest GFR values. Note that GAMLSS includes fractional polynomials, penalized splines and various other additive terms (see Stasinopoulos and Rigby (2007)).

5.4. Alternative Models

In this section we summarise the results of fitting some alternative models to the three datasets above, namely, (i) GAMLSS models of Rigby and Stasinopoulos (2005) as implemented in the `gamlss.cens` package in `R` (Stasinopoulos et al., 2018a), and (ii) penalized generalized survival (GS) models of Liu et al. (2018) implemented in the `rstpm2` package in `R` (Clements and Liu, 2019). Some selected "better-fiting" models appear in Table 13 with further details (including comparisons with Kaplan-Meier curves) deferred to the Supplementary Material.

Having fitted a wide variety of GAMLSS models to the three datasets, none improved significantly on the APGW-MPR models (see Supplementary Material). We should highlight that this is quite a large data-fitting exercise, and not one to be recommended as a general technique. This is not a criticism of GAMLSS, which is an impressive algorithmic framework offering many modelling choices, and note that the GAMLSS team do not necessarily recommend carrying out such an exercise. However, it is important to consider, among many potential modelling choices, which are likely to be most useful. We recommend the use of more "general purpose" distributions, and, in particular, advocate the APGW, along with knowledge of where to best place covariates, e.g., in Section 3, we suggest the use of a (horizontal or vertical) scale parameter, and a (power) shape parameter. In terms of model fit, only the generalized gamma emerges as a general competitor to the APGW across all three datasets, but this is expected based on Section 2.2 where we highlight some relative advantages of the APGW. Note from Table 13 that fitting the APGW is much less computationally intensive than the generalized gamma (likely due to the non-analytic form of survivor function in the latter), and, although the AIC/BIC values are slightly higher for the APGW models in the examples considered here, the model fit is similar (see Supplementary Material).

The GS models do not improve on the APGW model in the case of the lung cancer or

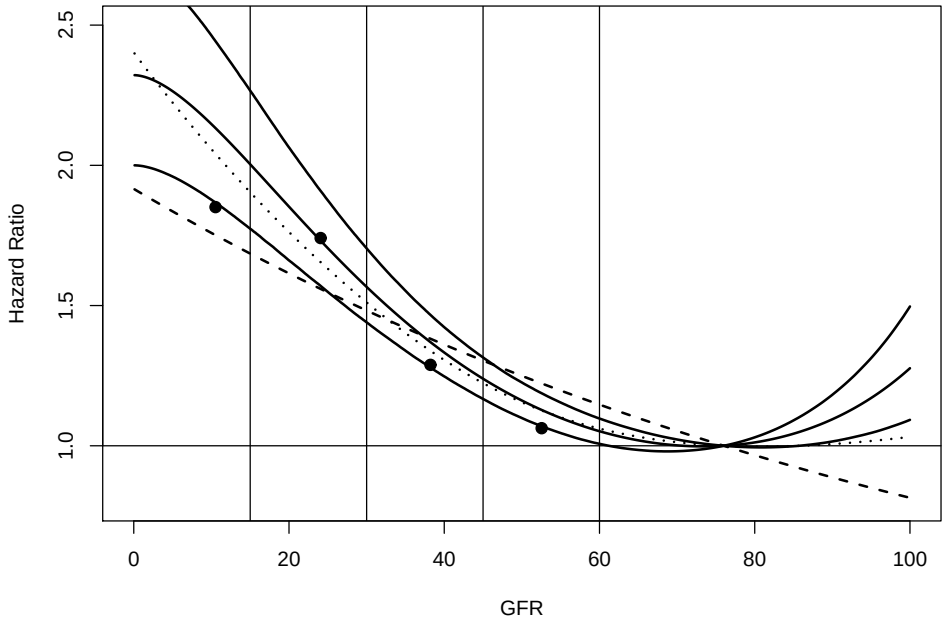

Fig. 7. Hazard ratios (relative to GFR $= 76$) for $M_F(\beta)$ with 95% confidence intervals (solid). Hazard ratios without confidence intervals are also shown for a Cox model with a penalized B-spline effect (dotted), $M_L(\beta)$ (dash), and $M(\beta)$ (points) where the points are located on the x-axis according to the mean GFR value in each of the GFR categories (and these categories are indicated using vertical lines). Line of equality (horizontal) also shown.

Table 13. Summary of models

Model	Lung Cancer ($n = 855$)			Melanoma ($n = 284$)			Kidney ($n = 6157$)		
	APGW $M(\beta, \alpha)$	GAMLSS `GG(2)`	GS PGST	APGW $M(\beta)$	GAMLSS `GG(1)`	GS PGS	APGW $M(\beta)$	GAMLSS `GG(1)`	GS PGS
$\dim(\theta)$	11	11	16.6	4	4	8.1	7	7	13.5
$\ell(\hat{\theta})$	-1927.0	-1924.9	-1923.6	-368.8	-365.9	-365.9	-6744.3	-6742.9	-6650.6
AIC	3876.0	3871.8	3880.3	745.6	739.9	748.0	13502.5	13499.8	13328.2
BIC	3928.2	3924.0	3959.1	760.2	754.5	777.7	13549.6	13546.9	13377.4
Time	0.57	1.92	501.46	0.06	1.03	0.32	3.00	170.00	30.46

$\ell(\hat{\theta})$, the value of the log-likelihood; $\dim(\theta)$, the dimension of the model, i.e., number of parameters (or effective number of parameters for the GS models); `GG(1)` and `GG(2)` are GAMLSS generalized gamma models where the first has covariate-dependent μ parameter and the second has covariate-dependent μ and σ parameters (and, here, "μ" and "σ" is GAMLSS notation); PGS, penalized generalized survival model; PGST, penalized generalized survival model with time-varying coefficients. Time, average time in seconds for estimation procedure to converge using an Intel® Core™ i5-6200U 2.3GHz processor. The APGW models were fitted to the data using the `nlm` function in `R` where the analytic score functions were provided (see Supplementary Material for `R` code).

melanoma datasets, but the improvement is significant in the case of the kidney data where the APGW model does not fit perfectly. Interestingly, in this latter case, the hazard ratios (not shown) are almost identical to those of the APGW model. Of course, this spline-based approach can adapt to any baseline in an essentially "non-parametric" way, but, clearly, this is not always required; the APGW model is a flexible fully parametric model, covering many important hazard shapes with minimal complexity. Note, in particular, from Table 13 that the penalized splines approach was quite computationally intensive in the case of the lung cancer data. This is due to the fact that time-varying coefficients were required necessitating two roughness penalties (one for the baseline distribution, and one for the time-varying effect).

6. Discussion

Our proposed APGW-MPR modelling framework is highly flexible and can adapt readily to a wide variety of applications in survival analysis and reliability. In particular, this framework includes the practically important AFT and PH models, and generalises them through shape regression components. Furthermore, the APGW baseline model covers the primary shapes of hazard function (constant, increasing, decreasing, up-then-down, down-then-up) within some of the most popular survival distributions (log-logistic, Burr type XII, Weibull, Gompertz) using only two shape parameters.

In practice, the full four-component APGW-MPR model is likely to be more flexible than is required for most purposes. In fact, we suggest that covariates should appear via just one scale-type component (τ or β), along with the α shape component which permits survivor functions with differing shapes and indicates departures from more basic AFT or PH effects, while $\nu = \log(\kappa + 1)$ is a covariate-independent parameter which allows us to choose among distributions within one unified framework. We have found that the scale-type parameters (τ and β) are highly intertwined in the sense that they cannot be estimated simultaneously within the same model reliably, and are highly correlated. This is true across the full range of distributions (varying ν), going well beyond the well-known Weibull case in which the two scale components are equivalent. The implication of this is that, in terms of performance gain, the movement from AFT to PH modelling (or vice versa) might not be very large, whereas we have found that modelling the shape is a more fruitful alteration to the regression specification.

Finally, the perspective of this paper has been to investigate survival modelling generally, to cover some of the most popular models, and to discover some of the better modelling choices that can be made within this framework. Although we have developed these ideas in a fully parametric context, non-parametric equivalents, while possible, are beyond the scope of the present paper (but are investigated in a separate line of work (Burke et al., 2019)). However, it is worth highlighting that perhaps too much emphasis is placed on non-/semi-parametric approaches in survival analysis whereby undue weight is attached to the flexibility of the baseline distribution in comparison to the flexibility of its regression structure. Our general approach to survival modelling provides a framework within which one can consider the most important components of survival modelling (including which might potentially be modelled non-parametrically), and we believe that this insight can lead to better modelling practice in general.

Acknowledgements

The first author would like to thank the Irish Research Council (www.research.ie) for supporting this work (New Foundations award). The authors would like to thank Prof. Austin Stack for

providing us with access to anonymized data from the National Kidney Disease Surveillance Programme (NKDSP), and the anonymous referees whose thoughtful comments led to an improved manuscript.

References

Anderson, J. and Senthilselvan, A. (1980). Smooth estimates for the hazard function. *J. R. Stat. Soc. B*, 42:322–327.

Bagdonaviçius, V. and Nikulin, M. (2002). *Accelerated Life Model; Modeling and Statistical Analysis*. Chapman & Hall/CRC, Boca Raton, FL.

Burke, K., Eriksson, F., and Pipper, C. B. (2019). Semiparametric multi-parameter regression survival modelling. *Scand. J. Stat.*, doi.org/10.1111/sjos.12416.

Burke, K. and MacKenzie, G. (2017). Multi-parameter regression survival modelling: An alternative to proportional hazards. *Biometrics*, 73:678–686.

Chao, C., Yubo, Z., Yingwei, P., and Jiajia, Z. (2012). *smcure: fit semiparametric mixture cure models*. R package version 2.0.

Chen, M., Ibrahim, J., and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *J. Am. Stat. Assoc.*, 94:909–919.

Chen, Y. and Jewell, N. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88:687–702.

Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Stat. Probabil. Lett.*, 49:155–161.

Clements, M. and Liu, X. (2019). *rstpm2: Generalized Survival Models*. R package version 1.4.5 (https://CRAN.R-project.org/package=rstpm2).

Cole, T. and Green, P. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat. Med.*, 11:1305–1319.

Cox, C. and Matheson, M. (2014). A comparison of the generalized gamma and exponentiated Weibull distributions. *Stat. Med.*, 33:3772–3780.

Cox, D. (1972). Regression models and life tables (with Discussion). *J. R. Stat. Soc. B*, 34:187–220.

Dimitrakopoulou, T., Adamidis, K., and Loukas, S. (2007). A lifetime distribution with an upside-down bathtub-shaped hazard function. *IEEE T. Reliab.*, 56:308–311.

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *J. Am. Stat. Assoc.*, 83:414–425.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Stat. Sci.*, 11:89–121.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on the new mode of determining the value of life contingencies. *Philos. T. R. Soc. A*, 115:513–580.

Hothorn, T., Möst, L., and Bühlmann, P. (2018). Most likely transformations. *Scand. J. Stat.*, 45:110–134.

Ibrahim, J., Chen, M., and Sinha, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, 57:383–388.

Jones, M. and Noufaily, A. (2015). Log-location-scale-log-concave distributions for survival and reliability analysis. *Electron. J. Stat.*, 9:2732–2750.

Jones, M., Noufaily, A., and Burke, K. (2020). A bivariate power generalized Weibull distribution: a flexible parametric model for survival analysis. *Stat. Methods Med. Res., to appear*.

Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *J. Am. Stat. Assoc.*, 53:457–481.

Kirkwood, J., Strawderman, M., Ernstoff, M., Smith, T., Borden, E., and Blum, R. (1996). Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the Eastern Cooperative Oncology Group Trial EST 1684. *J. Clin. Oncol.*, 14:7–17.

Kooperberg, C., Stone, C., and Truong, Y. (1995). Hazard regression. *J. Am. Stat. Assoc.*, 90:78–94.

Liu, X., Pawitan, Y., and Clements, M. (2018). Parametric and penalized generalized survival models. *Stat. Methods Med. Res.*, 27:1531–1546.

Matheson, M., Muñoz, A., and Cox, C. (2017). Describing the flexibility of the generalized gamma and related distributions. *J. Stat. Dist. Appl.*, 4:Article 15.

McLain, A. and Ghosh, S. (2013). Efficient sieve maximum likelihood estimation of time-transformation models. *J. Stat. Theory Pract.*, 7:285–303.

Nadarajah, S. and Haghighi, F. (2011). An extension of the exponential distribution. *Statistics*, 45:543–558.

Nikulin, M. and Haghighi, F. (2009). On the power generalized Weibull family: model for cancer censored data. *Metron*, 67:75–86.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Stat.*, 11:453–466.

Rigby, R. and Stasinopoulos, D. (1996). A semi-parametric additive model for variance heterogeneity. *Stat. Comput.*, 6:57–65.

Rigby, R. and Stasinopoulos, D. (2005). Generalized additive models for location, scale and shape. *J. R. Stat. Soc. C.*, 54:507–554.

Rosenberg, P. (1995). Hazard function estimation using B-splines. *Biometrics*, 51:874–887.

Royston, P. (2001). Flexible parametric alternatives to the Cox model, and more. *Stata J.*, 1:1–28.

Royston, P. and Altman, D. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J. R. Stat. Soc. C*, 43:429–453.

Royston, P. and Lambert, P. (2011). *Flexible parametric survival analysis using Stata: beyond the Cox model*. Stata Press.

Royston, P. and Parmar, M. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat. Med.*, 21:2175–2197.

Royston, P. and Sauerbrei, W. (2007). Improving the robustness of fractional polynomial models by preliminary covariate transformation: A pragmatic approach. *Comp. Stat. Data An.*, 51:4240–4253.

Rubio, F., Remontet, L., Jewell, N., and Belot, A. (2019). On a general structure for hazard-based regression models: An application to population-based cancer research. *Stat. Methods Med. Res.*, 28:2404–2417.

Sauerbrei, W., Royston, P., and Look, M. (2007). A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biometrical J.*, 49:453–473.

Stasinopoulos, D. and Rigby, R. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *J. Stat. Softw.*, 23:1–46.

Stasinopoulos, D. and Rigby, R. (2019). *gamlss: Generalised Additive Models for Location Scale and Shape*. R package version 5.1-3 (http://CRAN.R-project.org/package=gamlss).

Stasinopoulos, D., Rigby, R., and Akantziliotou, C. (2008). Instructions on how to use the gamlss package in R. Technical report, STORM Research Centre, London Metropolitan University, London. (http://www.gamlss.com).

Stasinopoulos, D., Rigby, R., and Mortan, N. (2018a). *gamlss.cens: Fitting an Interval Response Variable Using 'gamlss.family' Distributions*. R package version 5.0-1 (http://CRAN.R-project.org/package=gamlss.cens).

Stasinopoulos, M., Rigby, R., and De Bastiani, F. (2018b). GAMLSS: A distributional regression approach. *Stat. Model.*, 18:248–273.

Strasak, A., Umlauf, N., Pfeiffer, R., and Lang, S. (2011). Comparing penalized splines and fractional polynomials for flexible modelling of the effects of continuous predictor variables. *Comp. Stat. Data An.*, 55:1540–1551.

Tanner, M. and Wong, W. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Stat.*, 11:989–993.

Therneau, T. and Lumley, T. (2019). *survival: Survival Analysis*. R package version 2.44-1.1 (http://CRAN.R-project.org/package=survival).

Tsodikov, A., Ibrahim, J., and Yakovlev, A. (2003). Estimating cure rates from survival data: an

alternative to two-component mixture models. *J. Am. Stat. Assoc.*, 98:1063–1078.

Whittemore, A. and Keller, J. (1986). Survival estimation using splines. *Biometrics*, 42:495–506.

Xie, M., Tang, Y., and Goh, T. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliab. Eng. Syst. Safe.*, 73:279–285.

Younes, N. and Lachin, J. (1997). Link-based models for survival data with interval and continuous time censoring. *Biometrics*, 53:1199–1211.